Low back pain: Treatment, health effects, and costs

Berghuis-Mutubuki, E.N.

2021

document version
Publisher's PDF, also known as Version of record

Low Back Pain: Treatment, Health Effects, and Costs

Elizabeth
Nyasha Mutubuki

VU VRIJE UNIVERSITEIT AMSTERDAM

Low Back Pain: Treatment, Health Effects and Costs

Elizabeth Nyasha Mutubuki

The studies presented in this thesis were conducted at the Department of Health Sciences of the VU University, Amsterdam. This PhD thesis was embedded within Amsterdam Movement Sciences research institute, at the Department of Health Sciences, Vrije Universiteit Amsterdam.

The printing of this thesis was financially supported by the Scientific College Physical Therapy (WCF) of the Royal Dutch Society for Physical Therapy (KNGF), the VU University and the Amsterdam Movement Sciences research institute.

English title: Low Back Pain: Treatment, Health Effects and Costs
Dutch title: Lage rugpijn: Behandeling, Gezondheidseffecten en Kosten

Lay-out: Cindy van Rees
Cover: Cindy van Rees
Printed by: Blurb

VRIJE UNIVERSITEIT

Low Back Pain: Treatment, Health Effects and Costs

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor of Philosophy aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. V. Subramaniam,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Bètawetenschappen
op vrijdag 29 januari 2021 om 13.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door

Elizabeth Nyasha Mutubuki

geboren te Harare, Zimbabwe

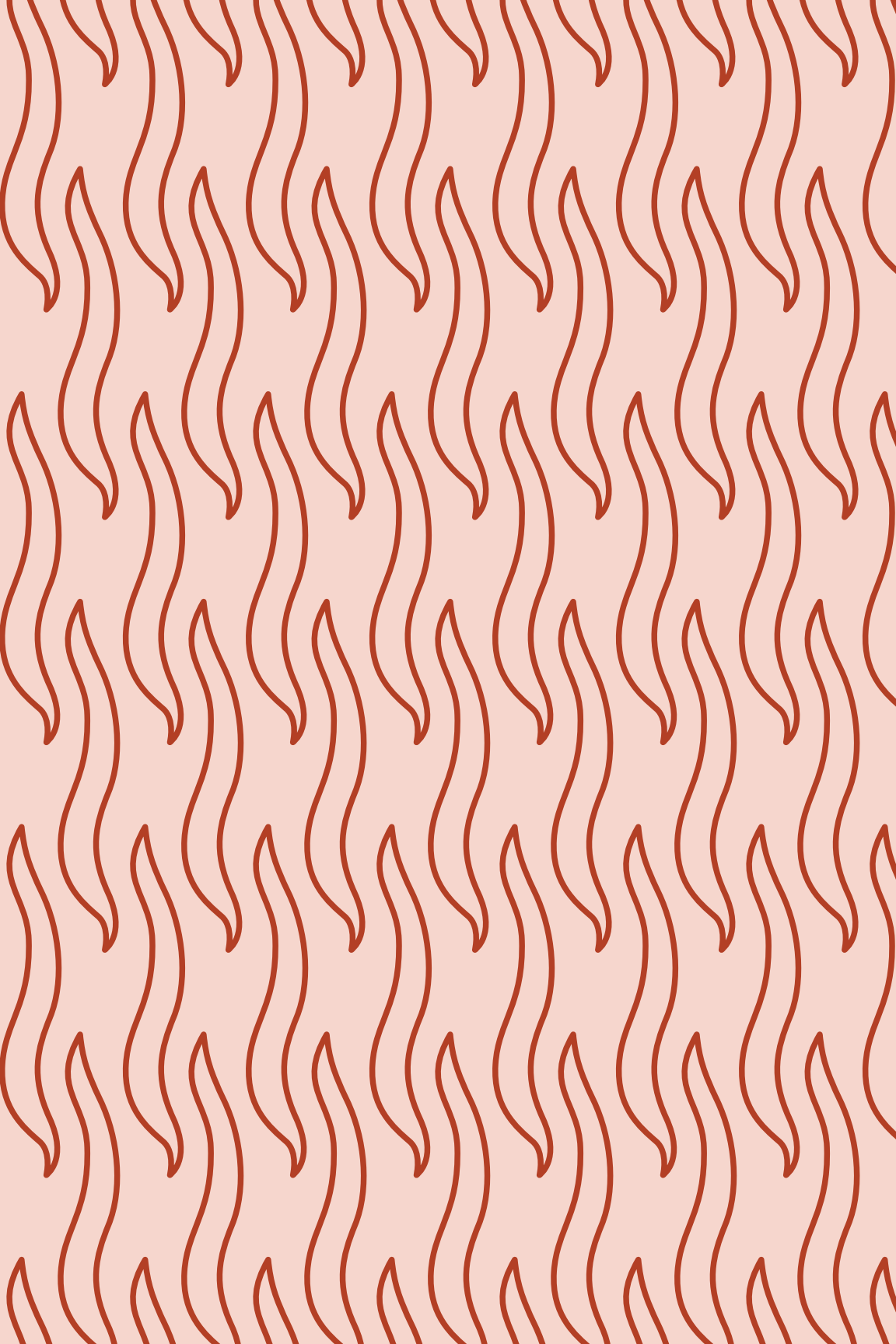
promotoren:	prof.dr. R.W.J.G. Ostelo
	prof.dr. M.W. van Tulder
copromotoren:	dr. J.M. van Dongen
	dr. C.L.A. Vleggeert-Lankamp

to Savannah and Aurora

More and more I have come to admire resilience.
Not the simple resistance of a pillow, whose foam
returns over and over to the same shape, but the
sinuous
tenacity of a tree: finding the light newly blocked
on one side,
it turns in another. A blind intelligence, true.
But out of such persistence arose turtles, rivers,
mitochondria, figs - all this resinous, unretractable
earth.

Jane Hirshfield, 'Optimism'

CONTENTS



General Introduction

1

General Introduction

GENERAL INTRODUCTION

1.1 Low back pain

Definition, prevalence, and incidence of low back pain

Low back pain (LBP) is not a disease, but a symptom[1]. It is commonly defined as pain, muscle tension or stiffness between the lower rib margins and the buttock creases with or without sciatica[1, 2]. LBP can also be defined as being specific or non-specific[3]. The aetiology of non-specific LBP is unknown. This group makes up about 80-90% of LBP patients[4]. In specific LBP, the cause of the symptoms is known and may include tumors, fractures, infection, and lumbar disc herniation. Specific LBP requires specific management, targeting the cause[1].

LBP can also be classified based on duration. When LBP symptoms persist for less than twelve weeks, it is classified as acute LBP, and chronic if symptoms persist for more than twelve weeks. The majority of people have one brief episode of acute LBP during their lifetime and most will recover within one year[2]. However, some will become chronic LBP patients with persisting or fluctuating pain with periods of no pain or heightened pain intensity[5]. LBP has also been described as an episodic condition and this has questioned the notion of acute versus chronic LBP[6]. This notion of acute versus chronic, presents LBP as unrelated acute episodes or chronic continuous pain[5, 6].

LBP is a very widespread health complaint, and is a burden to both society and patients. It is the leading cause of Years Lived with Disability (YLD) globally (Figure 1)[7]. According to the 2016 global burden of disease study, which assessed disease burden for 328 causes in 195 countries, LBP was in the top 10 causes of YLDs in 188 of the countries (Figure 1)[7]. In 2016, around 57 million YLDs were found to be associated with LBP and these have increased by more than 50% since 1990[7]. The high prevalence of LBP is partly responsible for its global burden[2]. Other factors that are responsible for the high global burden of LBP are population increase and aging[1].

Mean prevalence estimates of LBP are higher in high income countries compared to low and middle income countries[4]. The

global point prevalence of activity-limiting LBP was estimated at 7.3% in 2015. This implies that at that given moment around 540 million people worldwide were affected by LBP[8]. The lifetime prevalence of LBP is reported to range from 60 to 85%[9-12]. This indicates that people have a relatively high chance of developing an episode of LBP at any point in time during their life. In the Netherlands, an estimated 44% of adults will experience at least one episode of LBP during their life[13]. People who smoke, are overweight, have physically demanding jobs and mental and physical comorbidities have a higher risk of developing LBP compared to their counterparts[1]. Moreover, all ages can be affected by LBP[2], but the highest prevalence of LBP is reported between the age of 40-80 years, in women[4] and low socioeconomic status groups of society[1]. The prevalence of LBP is likely to increase in the upcoming decades due to aging of the population[14, 15].

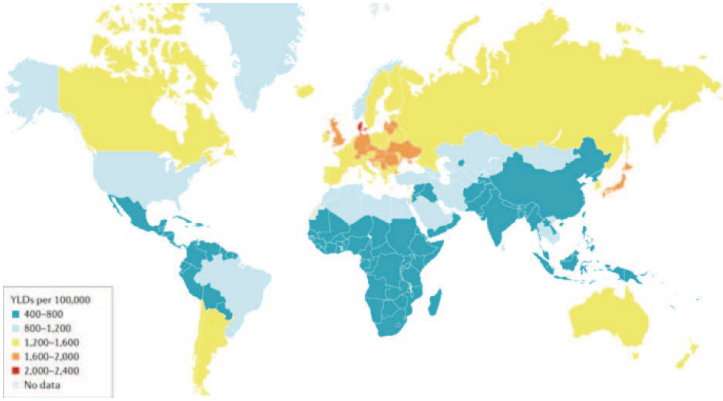


Figure 1: Years lived with disability for low back pain. Low back pain was one of the leading causes of years lived with disability (YLDs) in high- income, high- middle-income and middle-SDI (Socio- Demographic Index) quintile countries in 2016. Data from Global Burden of Disease Study, 2016[2].

Cost of low back pain

LBP is associated with high societal costs, of which the biggest share is due to increased absence from work and reduced productivity while being at work[14]. In the Netherlands, total societal costs from LBP were estimated to be as high as 3.5

billion euros in 2007, which equals about 0.6% of the Dutch gross national product[14]. In the United States, more than 100 billion Dollars was the estimated annual total societal cost of LBP[16, 17] and 9.0 billion Australian Dollars in Australia[18], 6.6 billion Euros in Switzerland[19], and 12.3 billion British Pounds in the UK[20]. The majority of LBP patients do not seek treatment; hence, it is very likely that the majority of the total societal costs from LBP stem from a relatively small group of chronic LBP patients[2, 21, 22]. As indicated above, absenteeism and a loss of productivity while being at work are the most important drivers of these societal costs in LBP[23]. Hence, interventions resulting in a decrease in disability and early return to work may result in the biggest cost-savings.

Sciatica

Sciatica, also known as lumbosacral radicular syndrome (LRS), is a specific kind of LBP commonly caused by a lumbar disc herniation. Sciatica is characterized by radiating lower limb pain into a particular dermatome, with or without sensory and or motor deficits[24]. The pathophysiology of sciatica is attributed to a complex interplay of inflammatory, immunological, and pressure related processes[25]. Nonetheless, the main source of pain is the impingement of a nerve root, which plays a role in about 85% of sciatica cases[26].

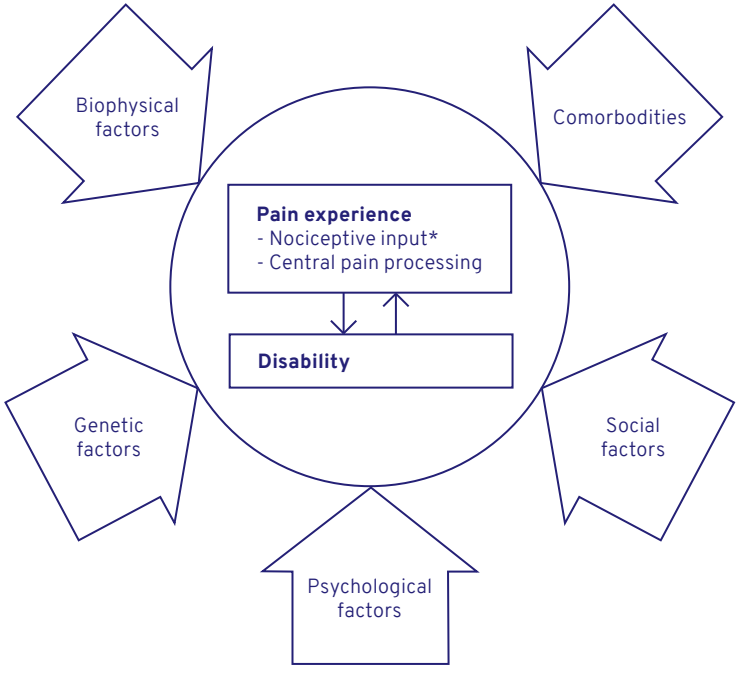
Sciatica is a prevalent health condition[2, 27] that is associated with high levels of pain and disability and low levels of health-related quality of life[28-30]. Life time prevalence estimates of sciatica vary from 12.2% to 43%, and its point prevalence ranges from 1.6% to 13.4%[31-36]. The incidence rate of sciatica in Western countries is estimated at 5 cases per 1,000 adults[37]. In the Netherlands, the incidence rate of sciatica is 9.4 cases per 1,000 patients per year[38] and there were 117,200 new cases of sciatica in 2017[39].

For most people with acute sciatica, prognosis is generally good[30, 40, 41], but about 30% of patients still have complaints after one year[42], and of those, about 10% will ultimately receive surgery[43]. The annual societal cost of sciatica is estimated to be €1.2 billion in the Netherlands[44].

1.2 Relationship between low back pain, outcomes and costs

Various factors contribute to persistent disabling LBP (Figure 2). These factors include genetic, psychological, social, biophysical as well as genetic factors, and comorbidities[1]. Hence, persisting disabling LBP does not only result from nociceptive input[1]. When LBP persists for a longer period it can result in higher functional disability, higher absenteeism, depression, insomnia, anxiety, poor health-related quality of life and high costs[2]. Understanding the interaction between these factors and knowing which of these factors predicts high costs can aid in the management of LBP.

Figure 2: Contributors to low back pain and disability
Model includes key contributors to low back pain and disability but does not show the interactions between the contributors[1].



Predictors of high societal costs among LBP patients

The economic burden of LBP on families, industry, governments, communities, and individuals is substantial[9, 45]. A proactive approach towards cost reductions requires identifying patients who are at high-risk of having high costs well before substantial avoidable costs have been incurred and health status has deteriorated[46]. Identifying predictors of having high societal costs among LBP patients provides an opportunity to explore ideas for initiatives or policy measures aimed at reducing and/or preventing those costs. Up till now, many prediction studies in LBP have investigated factors predicting whether acute LBP becomes chronic[47-49]. Although various predictive factors for chronicity in LBP have been identified, associations are typically weak and there has been limited success in using this information to manage or prevent chronic LBP[50, 51]. Other prediction studies in LBP have investigated predictive factors for return to work, disability and healthcare utilization. A few studies have explored predictive factors for high societal costs[52-55]. Hence, prediction models for having high societal costs in LBP are lacking. For this reason, predictive factors for high societal costs among chronic LBP patients will be identified in this thesis *(Chapter 3)*.

The relationship between pain/function and quality of life/costs

Persisting pain and limitations in function experienced by LBP patients can result in disability and poor health-related quality of life[2, 56-59], and are associated with higher costs[47-49, 51, 60, 61]. To illustrate, a study by Chiarotto found higher levels of pain severity to be correlated with lower levels of health-related quality of life in LBP patients[62]. Another study by Horng et al. reported significant correlations between pain intensity and disability and health-related quality of life[57]. Quality of life scores of patients with chronic LBP were even found to be comparable to those of individuals with a life-threatening diagnosis[63]. Moreover, LBP patients with higher levels of pain severity were found to be more likely to seek healthcare and to be absent from work[59]. However, studies evaluating these relationships have been cross-sectional in nature, meaning that they investigated whether pain severity and/or disability were associated with health related quality of life and/or healthcare

and societal costs at a certain point in time. Hence, they did not provide insight into whether individual changes in one variable (e.g., pain severity) are related to individual changes in another (e.g., health-related quality of life). These relationships, were both variables are measured and compared over time, can only be studied using a longitudinal study design[64]. The longitudinal relationships between pain severity and disability versus health-related quality of life, healthcare and societal costs among chronic LBP patients will therefore be explored in this thesis *(Chapter 2)*.

1.3 Effectiveness and cost-effectiveness of sciatica treatments

Wide variations in medical care for LBP in general, and sciatica in particular, exist despite advances in treatments[65]. In addition, there is a lack of clarity in the mechanism of action of many treatments for LBP and sciatica, and effect sizes of most treatments are small[66]. Consequently, firm conclusions regarding the effectiveness of such treatments cannot be made. Frequently provided treatments for sciatica include surgery, medication, exercises, advice to stay active, and injections[30, 67], all of which aim to reduce pain by either analgesics or through reduction of nerve root pressure. Effective and cost-effective treatments in sciatica will ensure a better management of sciatica and potential cost savings. Below, exercise and surgery for sciatica are further described.

Exercise

Exercise as therapy aims to improve function, reduce pain, and hence speed-up recovery among sciatica patients[68]. Benefits of exercise include improved muscle strength, cardiorespiratory, and cardiovascular function[68, 69]. In clinical settings, consensus concerning the benefits and most optimal type of exercise for patients with sciatica is still lacking. This is evidenced by the fact that group exercise programs, such as aerobic exercise, mind–body or a combination of approaches, are recommended by the UK NICE guideline, while the Dutch general practitioner guideline discourages routine referrals for exercise and recommends limiting demanding activities of

daily living[34, 70, 71]. The Danish multidisciplinary guidelines recommend considering supervised exercises as an addition to usual care[72]. Supervised exercise therapy includes directional exercises, motor control exercise, nerve mobilization, or strength exercises, but no specific recommendations for any specific type of exercise treatment are made[72]. The goal of exercise differs and includes improvement of range of motion, strength, and core stability. Muscles trained also differ and include gluteus, leg muscles, lower back and core muscles. Some exercises are given as adjunct therapy to e.g. injections. Many types of exercise are available, and systematic reviews might help in making choices for the most optimal exercise treatment in sciatica patients. Hence, a systematic review to determine the effectiveness of exercise therapy on sciatica will be presented in this thesis *(Chapter 6)*.

Surgery

Surgery is recommended in sciatica patients if pain persists following conservative management[72, 73]. Micro-discectomy is often performed in hernia patients, with the aim to remove the symptomatic disc herniation, by a minimal unilateral transflaval approach with magnification, and the patient under general or spinal anaesthesia. About 12,000 sciatica surgeries are performed per year in the Netherlands. The healthcare costs associated with early surgery are acceptable and compensated by the difference in work absenteeism[74]. Surgery can result in complications, such as infection, nerve root damage, and residual complaints, though they rarely occur[75]. In the Netherlands, satisfactory results are gained in over 90% in the first period after surgery, but in the follow-up recovery is estimated to be present in 69-79% of patients two years after receiving surgery. Repeated surgery is reported to occur in about 10-15% of the patients, mostly due to recurrent disc herniation at the same level[76]. However, those figures come mostly from controlled trials which are not always representative of daily practice. In the long–term, surgical and non-surgical management of lumbar hernia are reported to be equally successful[77]. Surgery enables a faster recovery, relief from leg pain and earlier return to normal activities[78]. In addition, it is associated with some side effects; hence, preventing surgery might still be worthwhile.

A possible means to accomplish this is to offer sciatica patients a "last alternative" while being on the waiting list for surgery. In a pilot study, a combination therapy consisting of mechanical diagnosis and treatment (MDT) and transforaminal epidural steroid injections (TESIs) was reported to have the potential to reduce the amount of lumbar hernia surgeries[79]. The pilot study, however, lacked a control group, and randomization. Moreover, the cost-effectiveness of the intervention compared with usual care was not assessed. For that reason, an evaluation of the effectiveness and cost-effectiveness of combination therapy compared to usual care after a follow-up of 6 months is included in this thesis *(Chapter 4 and 5)*.

1.4 Methodological studies

Given the economic burden of LBP, it is important to use measurement instruments and apply treatments that are based on high quality evidence. That is, instruments should measure what they set out to measure and evidence on the (cost-) effectiveness of LBP treatments should be based on evidence that is not biased by poor methodology. Methodologically sound evidence and high quality trials will provide good information regarding efficient usage of resources and potential savings.

Measuring recovery in sciatica

Recovery rates of LBP in literature vary wildly due to differences in definitions and measurement instruments used for recovery, hindering comparisons. Although not part of the core outcomes set for LBP, the GPE scale is often used to measure self-perceived recovery in LBP patients[80-82]. The GPE scale requires patients to indicate their improvement or deterioration since a given time point[83]. It is favored because it is easy to use, quick to administer and its test-retest reliability is excellent[83]. However, there are construct validity concerns regarding the GPE scale. The most important concern is whether it is a true "transition scale". That is, GPE scores are reported to be significantly prejudiced by a patient's current health status. In some studies this effect became more obvious with increasing time periods[83]. However, in their study, Kamper et al performed a complete-case analysis, hence their conclusions

could have been influenced by selective drop-out of patients. Moreover, possible confounding factors, such as age and duration of complaints, were not controlled for and Kamper et. al only included patients in their study who suffered substantial residual complaints six weeks after lumbar disc surgery, and not all postoperative sciatica patients[84]. For these reasons, the construct validity of the GPE in postoperative sciatica patients will be explored in this thesis, while considering the shortcomings of the previous studies on this topic *(Chapter 7)*.

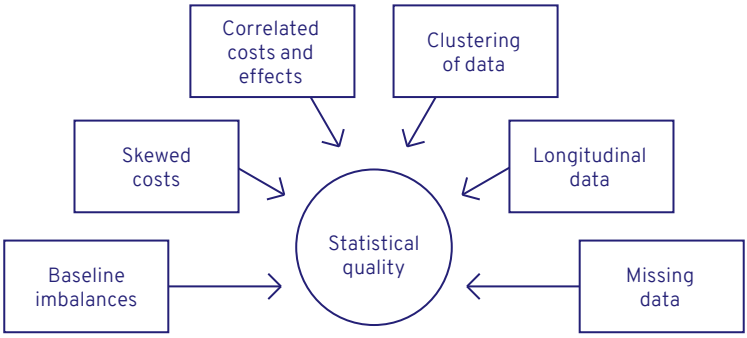
Economic evaluation in low back pain

Healthcare resources are scarce; hence, it is of paramount importance to utilize available resources efficiently. Therefore, decisions about the implementation and/or reimbursement of LBP treatments should not only be based on their effectiveness, but also on their so-called cost-effectiveness. Economic evaluations aim to provide information on the cost-effectiveness of interventions by assessing whether the additional health effects of a new intervention justify its additional costs as compared to an alternative intervention[85]. Currently, however, high quality economic evaluations are lacking for a broad range of LBP treatments, therefore robust conclusions regarding the cost-effectiveness of such treatments cannot be made[86]. To inform healthcare decision-makers about effective and cost-effective interventions in LBP it is essential that methodologically sound evidence and high quality trials and economic evaluations exist.

Great improvements in conducting and reporting of economic evaluations alongside clinical trials have been made in previous years[87], but the quality of the applied statistical methods remains far from optimal[87-89]. That is, baseline imbalances, skewed costs, the correlation between costs and effects, and missing data are not often adequately accounted for in trial-based economic evaluations[90], whereas they are part of the most important statistical challenges to trial-based economic evaluations as defined by van Dongen et al. (2019)(Figure 3). Failure to adequately account for these statistical issues in analyzing trial-based economic evaluations is worrisome, because use of insufficient statistical methods may lead to biased results, and as a result, invalid decisions causing wastage

of scarce resources[91]. In this thesis the impact of accounting and not accounting for baseline imbalances, skewed costs, correlated costs and effects, and missing data on results of trial-based economic evaluation will therefore be explored *(Chapter 8)*.

Figure 3. Statistical challenges to trial-based economic evaluations [90]



1.5 Aims and outline of this thesis

Objectives of this thesis

The overall aim of this thesis is to contribute to the development of a sound evidence base on:

1) the relationship between LBP, outcomes and costs *(Chapter 2 and 3)*
2) the effectiveness and cost-effectiveness of sciatica treatments *(Chapter 4, 5 and 6)*.

And to improve:

3) scientific methods in LBP research *(Chapter 7 and 8)*

Outline of this thesis:

This thesis consists of three themes (A-C), a general introduction, a general discussion and an English summary. Theme A will explore the relationships between LBP, outcomes and costs. Theme B will investigate the effectiveness and cost-effectiveness of treatments in sciatica. Theme C will look into the aforementioned methodological issues.

Theme A: Relationship between low back pain, outcomes and costs

Included in this theme is a prediction model and a longitudinal study to identify predictors of high costs and to explore the relationship between pain, disability, health-related quality of life and costs over time, respectively.

Overall research questions of this theme are:

1. What is the association between pain severity/disability with health-related quality of life and costs within and between individuals over a 3-month period? *(Chapter 2)*
2. Which factors predict high societal costs among chronic LBP patients? *(Chapter 3)*

Theme B: Effectiveness and cost-effectiveness of sciatica treatments

Included in this theme is a protocol, an evaluation of the effectiveness and cost-effectiveness of combination therapy compared to usual care after a follow-up of 6 months, and a systematic review on the effectiveness of exercise therapy in sciatica patients.

Overall research questions of this theme are:

1. Is combination therapy (MDT &TESIs) effective and cost-effective compared to usual care among sciatica patients with an indication for surgery? *(Chapter 4 and 5)*
2. Is exercise therapy for sciatica effective? *(Chapter 6)*

Theme C: Methodological studies

Included in this theme are two methodological papers. One paper explores the construct validity of the GPE. The other paper explores the impact on results of correcting for statistical challenges (i.e. baseline imbalances, skewed costs, correlation between costs and effects, missing data) in trial based economic evaluations.

Overall research questions of this theme are:

1. To what extend is the association between GPE and change in pain and functional status influenced by current health status? *(Chapter 7)*
2. Does the statistical approach in trial-based economic evaluations matter? *(Chapter 8).*

In the general discussion, *Chapter 9*, the main findings and methodological issues pertaining to the various chapters will be discussed, as well as their implications for research and practise.

REFERENCES

1. Hartvigsen J, Hancock MJ, Kongsted A, Louw Q, Ferreira ML, Genevay S, Hoy D, Karppinen J, Pransky G, Sieper J, Smeets RJ, Underwood M and Lancet Low Back Pain Series Working Group. What low back pain is and why we need to pay attention. *Lancet*. 2018.

2. Vlaeyen JWS, Maher CG, Wiech K, Van Zundert J, Meloto CB, Diatchenko L, Battié MC, Goossens M, Koes B, Linton SJ. Low back pain. *Nat Rev Dis Primers*. 2018; 4(1):52.

3. Itz CJ, Willems PC, Zeilstra DJ, and Huygen FJ. Dutch multidisciplinary guideline for invasive treatment of pain syndromes of the lumbosacral spine. *Pain Practice*. 2016;16(1):90-110.

4. Hoy D, Bain C, Williams G, March L, Brooks P, Blyth F, Woolf A, Vos T, Buchbinder R. A systematic review of the global prevalence of low back pain. *Arthritis Rheum*. 2012;64(6):2028-37.

5. Kongsted A, Kent P, Axen I, Downie AS, Dunn KM. What have we learned from ten years of trajectory research in low back pain? *BMC Musculoskelet Disord*. 2016;17:220.

6. Von Korff M, Saunders K. The course of back pain in primary care. *Spine*. 1996;21(24):2833-7

7. Vos T, Abajobir AA, Abate KH, Abbafati C, Abbas KM, Abd-Allah F, Abdulle AM, Abebo TA, Abera SF, Aboyans V, Abu-Raddad LJ et al. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet*. 2017;390(10100):1211-59.

8. Global Burden of Disease II, Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with

disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet*. 2016;388:1545–602.

9. Hoy D, March L, Brooks P, Woolf A, Blyth F, Vos T, Buchbinder R. Measuring the global burden of low back pain. *Best practice & research Clinical rheumatology*. 2010;24(2):155-65.

10. Andersson GB. Epidemiological features of chronic low-back pain. *The Lancet*. 1999;354(9178):581-5.

11. Krismer M, Van Tulder M. Low back pain (non-specific). *Best practice & research clinical rheumatology*. 2007;21(1):77-91.

12. Manchikanti L, Singh V, Datta S, Cohen SP, Hirsch JA. Comprehensive review of epidemiology, scope, and impact of spinal pain. *Pain physician*. 2009;12(4):E35-70.

13. Picavet HS, Schouten JS. Musculoskeletal pain in the Netherlands: prevalences, consequences and risk groups, the DMC3-study. *Pain*. 2003;102:167-78.

14. Lambeek LC, van Tulder MW, Swinkels IC, Koppes LL, Anema JR, van Mechelen W. The trend in total cost of back pain in The Netherlands in the period 2002 to 2007. *Spine*. 2011;36(13):1050-8.

15. Smith M, Davis, MA, Stano M, & Whedon JM. Aging baby boomers and the rising cost of chronic back pain: secular trend analysis of longitudinal Medical Expenditures Panel Survey data for years 2000 to 2007. *Journal of manipulative and physiological therapeutics*. 2013;36(1):2-11.

16. Dieleman JL, Baral R, Birger M, Bui AL, Bulchis A, Chapin A, et al. US spending on personal health care and public health, 1996-2013. *Jama*. 2016;316(24):2627-46.

17. Crow WT, & Willis DR. Estimating cost of care for patients with acute low back pain: a retrospective review of patient records. *The Journal of the American Osteopathic Association*. 2009;109(4):

229-33.

18. Walker BF, Muller R, Grant WD. Low back pain in Australian adults: the economic burden. *Pac J Public Health*. 2003;15(2):79-87.

19. Wieser S, Horisberger B, Schmidhauser S, Eisenring C, Brugger U, Ruckstuhl A, Dietrich J, Mannion AF, Elfering A, Tamcan O, Muller U. Cost of low back pain in Switzerland in 2005. *Eur J Health Econ*. 2011;12(5):455-67.

20. Maniadakis N, Gray A. The economic burden of back pain in the UK. *Pain*. 2000;84(1):95-103.

21. Ferreira ML, Machado G, Latimer J, Maher C, Ferreira PH, Smeets RJ. Factors defining care seeking in low back pain–A meta analysis of population based surveys. *Eur J Pain*. 2010;14(7):741-7.

22. Engel CC, von Korff M, Katon WJ. Back pain in primary care: predictors of high health-care costs. *Pain*. 1996;65(3):197-204.

23. Tsuboi Y, Murata S, Naruse F, & Ono R. Association between pain related fear and presenteeism among eldercare workers with low back pain. *Eur J Pain*. 2019;23(3):495-502.

24. Oosterhuis T, Costa LO, Maher CG, de Vet HC, van Tulder MW, Ostelo RW. Rehabilitation after lumbar disc surgery. *Cochrane Database Syst Rev*. 2014(3).

25. Stafford MA, Peng P, Hill DA. Sciatica: a review of history, epidemiology, pathogenesis, and the role of epidural steroid injection in management. *British Journal of Anaesthesia*. 2007;99(4):461-73.

26. Porchet F, Wietlisbach V, Burnand B, Daeppen K, Villemure JG, & Vader JP. Relationship between severity of lumbar disc disease and disability scores in sciatica patients. *Neurosurgery*. 2002;50(6):1253-60.

27. Maslak JP, Jenkins TJ, Weiner JA, Kannan AS, McCarthy MH, Hsu WK, Patel AA. Burden of Sciatica on US Medicare Recipients. *J Am Acad Orthop Surg*. 2019.

28. Konstantinou K, Hider SL, Jordan JL, Lewis M, Dunn KM, Hay EM. The impact of low back-related leg pain on outcomes as compared with low back pain alone: a systematic review of the literature. *Clin J Pain*. 2013;29(7):644-54.

29. Kristman VL, Hartvigsen J, Leboeuf-Yde C, Kyvik KO, Cassidy JD. Does radiating spinal pain determine future work disability? A retrospective cohort study of 22,952 Danish twins. *Spine*. 2012;37(11):1003-13.

30. Koes BW, van Tulder MW, Peul WC. Diagnosis and treatment of sciatica. *BMJ*. 2007;334(7607):1313-7.

31. Heliovaara M, Impivaara O, Sievers K, Melkas T, Knekt P, Korpi J, and Aromaa A. Lumbar disc syndrome in Finland. *J Epidemiol Community Health*. 1987;41(3):251-8.

32. Konstantinou K, Hider SL, Jordan JL, Lewis M, Dunn KM, Hay EM. The impact of low back-related leg pain on outcomes as compared with low back pain alone: a systematic review of the literature. *Clin J Pain*. 2013;29(7):644-54.

33. Oosterhuis T, Smaardijk V, Kuijer P, Hoving J, Frings-Dresen M. Actualisatie wetenschappelijke kennis en inzichten over het lumbosacraal radiculair syndroom. [Internet] [Report]. Uitvoeringsinstituut werknemersverzekeringen UWV; 2016 [updated April 2019. Dec 2016:[Available from: https://www.uwv.nl/overuwv/Images/Actualisatie-lumbosacraal-radiculair-syndroom.pdf.

34. Schaafstra A, Spinnewijn WEM, Bons S, Borg M, Koes BW, Ostelo R, Spijker-Huiges A, Burgers JS, Bouma M, Verburg AFE. NHG-Standaard Lumbosacraal radiculair syndroom (tweede herziening). *Huisarts & Wetenschap*. 2015;58(6):308-20.

35. Spijker-Huiges A, Groenhof F, Winters JC, van Wijhe M,

Groenier KH, van der Meer K. Radiating low back pain in general practice: incidence, prevalence, diagnosis, and long-term clinical course of illness. *Scand J Prim Health Care*. 2015;33(1):27-32.

36. Nielen MMJ, Boersma-van Dam ME, Schermer TRJ. Incidentie en prevalentie van gezondheidsproblemen in de Nederlandse huisartsenpraktijk in 2017. From: NIVEL Zorgregistraties eerste lijn [internet]. 2019 [cited 2019 April 16] 2019 [Available from: www.nivel.nl/nl/zorgregistraties-eerste-lijn/incidenties-en-prevalenties.

37. Cherkin DC, Deyo, RA, Loeser JD, Bush T, Waddell G. An international comparison of back surgery rates. *Spine*. 1994;19:1201-6.

38. Spijker-Huiges A, Groenhof F, Winters JC, van Wijhe M, Groenier KH, van der Meer K. Radiating low back pain in general practice: incidence, prevalence, diagnosis, and long-term clinical course of illness. *Scandinavian Journal of Primary Health Care*. 2015;33(1):27-32.

39. Nivel Zorgregistraties eerste lijn. Nivel Zorgregistraties. 2018.

40. Ropper AH, Zafonte RD. Sciatica. *N Engl J Med*. 2015;372(13):1240-8.

41. Vroomen PC, de Krom MC, Wilmink JT, Kester AD, Knottnerus JA. Lack of effectiveness of bed rest for sciatica. *N Engl J Med*. 1999 (340):418-23.

42. Vroomen PC, de Krom MC, Slofstra PD, Knottnerus JA. Conservative treatment of sciatica: a systematic review. *J Spinal Disord*. 2000;13(6):463-9.

43. van Beek E, Lemmens K, van Schooten G, Vlieger EJ. Reduceren van praktijkvariatie: budgettaire effecten van scherpere indicatiestelling. Reduceren van praktijkvariatie: budgettaire effecten van scherpere indicatiestelling. Breukelen: Plexus. 2010.

44. The Health Council of the Netherlands. Management of the lumbosacral radicular syndrome (sciatica) 1999. [Available from: https://www.healthcouncil.nl/documents/advisory-reports/1999/11/16/management-of-the-lumbosacral-radicular-syndrome-sciatica]

45. van Tulder M, Koes B, Bombardier C. Low back pain. *Best Pract Res Clin Rheumatol*. 2002;16(5):761-75.

46. Mutubuki EN, Luitjens MA, Maas ET, Huygen FJ, Ostelo RW, van Tulder MW, and van Dongen JM. Predictive factors of high societal costs among chronic low back pain patients. *European Journal of Pain*. 2019.

47. Klenerman L, Slade P, Stanley I, Pennie B, Reilly J, Atchison L, Troup JD, Rose MJ. The prediction of chronicity in patients with an acute attack of low back pain in a general practice setting. *Spine*. 1995;20(4):478-84.

48. Linton SJ, Halldén K. Can we screen for problematic back pain? A screening questionnaire for predicting outcome in acute and subacute back pain. *Clin J Pain*. 1998;14(3):209-15.

49. Valat JP, Goupille P, Védere V. Low back pain: risk factors for chronicity. *Rev Rhum Engl Ed*. 1997;64(3):189-94.

50. van Tulder M, Koes B, Bombardier C. Low back pain. *Best Pract Res Clin Rheumatol*. 2002;16(5):761-75.

51. Waddell G. Preventing incapacity in people with musculoskeletal disorders. *Br Med Bull*. 2006;77(1):55-69.

52. Becker A, Heiko H, Redaelli M, Strauch K, Chenot JF, Leonhardt C, Keller S, Baum E, Pfingsten M, Hildebrandt J, Basler H, Kochen MM, Donner-Banzhoff N. Low back pain in primary care: costs of care and prediction of future health care utilization. *Spine*. 2010;35(18):1714-20.

53. Lancourt J, Kettelhut M. Predicting return to work for lower

back pain patients receiving worker's compensation. *Spine*. 1992;17(6):629-40.

54. Pincus T, Burton AK, Vogel S, Field AP. A systematic review of psychological factors as predictors of chronicity/ disability in prospective cohorts of low back pain. *Spine*. 2002;27(5):E109-E20.

55. Skargren EI, Öberg BE. Predictive factors for 1-year outcome of low-back and neck pain in patients treated in primary care: comparison between the treatment strategies chiropractic and physiotherapy. *Pain*. 1998;77(2):2017.

56. Sadosky AB, Taylor-Stokes G, Lobosco S, Pike J, and Ross E. Relationship between self-reported low-back pain severity and other patient-reported outcomes: results from an observational study. *Clinical Spine Surgery*. 2013;26(1):8-14.

57. Horng YS, Hwang YH, Wu HC, Liang HW, Mhe YJ, Twu FC and Wang JD. Predicting health-related quality of life in patients with low back pain. *Spine*. 2005;30(5):551-5.

58. Von Korff M, Crane P, Lane M, Miglioretti DL, Simon G, Saunders K, Stang P, Nancy B and Kessler R. Chronic spinal pain and physical–mental comorbidity in the United States: results from the national comorbidity survey replication. *Pain*. 2005;113(3):331-9.

59. Sadosky AB, DiBonaventura M, Cappelleri JC, Ebata N, Fujii K. The association between lower back pain and health status, work productivity, and health care resource use in Japan. *Journal of pain research*. 2015;8:119.

60. Verkerk K, Luijsterburg PA, Heymans MW, Ronchetti I, Pool-Goudzwaard AL, Miedema HS and Koes BW. Prognosis and course of disability in patients with chronic nonspecific low back pain: a 5-and 12-month follow-up cohort study. *Phys Ther*. 2013;93(12):1603-14.

61. Buchbinder R, Blyth FM, March LM, Brooks P, Woolf AD, Hoy

DG. Placing the global burden of low back pain in context. *Best Pract Res Clin Rheumatol*. 2013;27(5):575-89.

62. Chiarotto A, Maxwell LJ, Ostelo RW, Boers M, Tugwell P and Terwee CB. Measurement properties of Visual Analogue Scale, Numeric Rating Scale and Pain Severity subscale of the Brief Pain Inventory in Patients With Low Back Pain: A Systematic Review. *The Journal of Pain*. 2019;20(3):245-263.

63. Fredheim OMS, Kaasa S, Fayers P, Saltnes T, Jordhøy M and Borchgrevink PC. Chronic non- malignant pain patients report poor health related quality of life as palliative cancer patients. *Acta Anaesthesiologica Scandinavica*, 2008;52(1):143-148.

64. Twisk JWR. Applied Longitudinal Data Analysis for Epidemiology. Cambridge: Cambridge University Press. 2013.

65. Deyo RA and Tsui-Wu YJ. Descriptive epidemiology of low-back pain and its related medical care in the United States. *Spine* (Phila Pa 1976), 1987;12(3):264-8.

66. Balagué F, Mannion AF, Pellisé F and Cedraschi C. Nonspecific low back pain. *The Lancet*, 2012;379(9814):482-491.

67. Furlan AD, Malmivaara A, Chou R, Maher CG, Deyo RA, Schoene M, Bronfort G, van Tulder MW and Editorial Board of the Cochrane Back, Neck Group. Updated Method Guideline for Systematic Reviews in the Cochrane Back and Neck Group. *Spine*, 2015;40(21):1660-1673.

68. Hayden JA, Cartwright J, van Tulder MW, and Malmivaara A. Exercise therapy for chronic low back pain. *Cochrane Database of Systematic Reviews*. 2012.

69. K. E. Powell, A. E. Paluch, and S. N. Blair. Physical Activity for Health: What Kind? How Much? How Intense? On Top of What? *Annual Review of Public Health*. 2011;32(1):349-365.

70. NICE, guideline, and [NG59]. “Low back pain and sciatica in over 16s: assessment and management.” https://www.nice.

org.uk/guidance/NG59/chapter/Recommendations#non-invasivetreatments-for-low-back-pain-and-sciatica (accessed 29 March 2019.)

71. National Institute for Health and Care Excellence. Low back pain and sciatica: management of non-specific low back pain and sciatica. *National Institute for health Care Excellence*, London, 2016.

72. Stochkendahl MJ, Hartvigsen J, Kongsted A, Aaboe J, Andersen M, Andersen MØ, Fournier G, Højgaard B, Jensen MB, Jensen LD, Karbo T, Kirkeskov L, Melbye M, Morsel-Carlsen L, Nordsteen J, Palsson TS, Rasti Z, Silbye PF, Steiness MZ, Tarp S and Vaagholt M. National Clinical Guidelines for non-surgical treatment of patients with recent onset low back pain or lumbar radiculopathy. *Eur Spine J*. 2018(1):60-75.

73. Schaafstra A, Spinnewijn WEM, Bons SCS, Borg MAJP, Koes BW, Ostelo RWJG, Spijker-Huiges A, Burgers JS, Bouma M, Verburg AFE. NHG-Standaard Lumbosacraal radiculair syndroom (tweede herziening). Huisarts Wet 2015;58(6):308-20.

74. van den Hout WB, Peul WC, Koes BW, Brand R, Kievit J, Thomeer RT; Leiden-The Hague Spine Intervention Prognostic Study Group. Prolonged conservative care versus early surgery in patients with sciatica from lumbar disc herniation: cost utility analysis alongside a randomised controlled trial. *BMJ*. 2008;336(7657):1351-4.

75. Awad JN, Moskovich R. Lumbar disc herniations: surgical versus nonsurgical treatment. *Clinical Orthopaedics and Related Research*, 2006;443:183-197.

76. Arts MP, Brand R, van den Akker ME, Koes BW, Bartels RH, Tan WF, Peul WC. Tubular diskectomy vs conventional microdiskectomy for the treatment of lumbar disk herniation: 2-year results of a double-blind randomized controlled trial. *Neurosurgery*. 2011;69(1):135-44; discussion 144.

77. Peul WC, van Houwelingen HC, van den Hout WB, Brand R,

Eekhof JA, Tans JT, Thomeer RT, Koes BW; Leiden-The Hague Spine Intervention Prognostic Study Group. Surgery versus prolonged conservative treatment for sciatica. *N Engl J Med*. 2007;356(22):2245-56.

78. Peul WC, van den Hout WB, Brand R, Thomeer RT, Koes BW and Leiden-The Hague Spine Intervention Prognostic Study Group. Prolonged conservative care versus early surgery in patients with sciatica caused by lumbar disc herniation: two year results of a randomised controlled trial. *BMJ.* 2008;336(7657):1355-1358.

79. van Helvoirt H, Apeldoorn AT, Ostelo RW, Knol DL, Arts MP, Kamper SJ, van Tulder MW. Transforaminal epidural steroid injections followed by mechanical diagnosis and therapy to prevent surgery for lumbar disc herniation. *Pain Medicine*, 2014;15(7):1100-1108.

80. Pengel LH, Refshauge KM, Maher CG. Responsiveness of pain, disability, and physical impairment outcomes in patients with low back pain. *Spine* (Phila Pa 1976). 2004;29(8):879-83.

81. van der Roer N, Ostelo RW, Bekkering GE, van Tulder MW, de Vet HC. Minimal clinically important change for pain intensity, functional status, and general health status in patients with nonspecific low back pain. *Spine* (Phila Pa 1976). 2006;31(5):578-82.

82. Stewart M, Maher CG, Refshauge KM, Bogduk N, Nicholas M. Responsiveness of pain and disability measures for chronic whiplash. *Spine* (Phila Pa 1976). 2007;32(5):580-5.

83. Kamper SJ, Ostelo RW, Knol DL, Maher CG, de Vet HC and Hancock MJ. Global Perceived Effect scales provided reliable assessments of health transition in people with musculoskeletal disorders, but ratings are strongly influenced by current status. *J Clin Epidemiol.* 2010;63(7):760-766.e1.

84. Ostelo RW, de Vet HC, Berfelo MW, Kerckhoffs MR, Vlaeyen JW, Wolters PM, van den Brandt PA. Effectiveness of behavioral

graded activity after first-time lumbar disc surgery: short term results of a randomized controlled trial. *Eur Spine J*. 2003;12(6):637-44.

85. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL and Torrance GW. Methods for the economic evaluation of healthcare programmes. *Oxford university press*, 2015.

86. van Dongen JM, Ketheswaran J, Tordrup D, Ostelo RWJG, Bertollini R and van Tulder MW. Health economic evidence gaps and methodological constraints in low back pain and neck pain: Results of the Research Agenda for Health Economic Evaluation (RAHEE) project. *Best Pract Res Clin Rheumatol*. 2016;30(6):981-993.

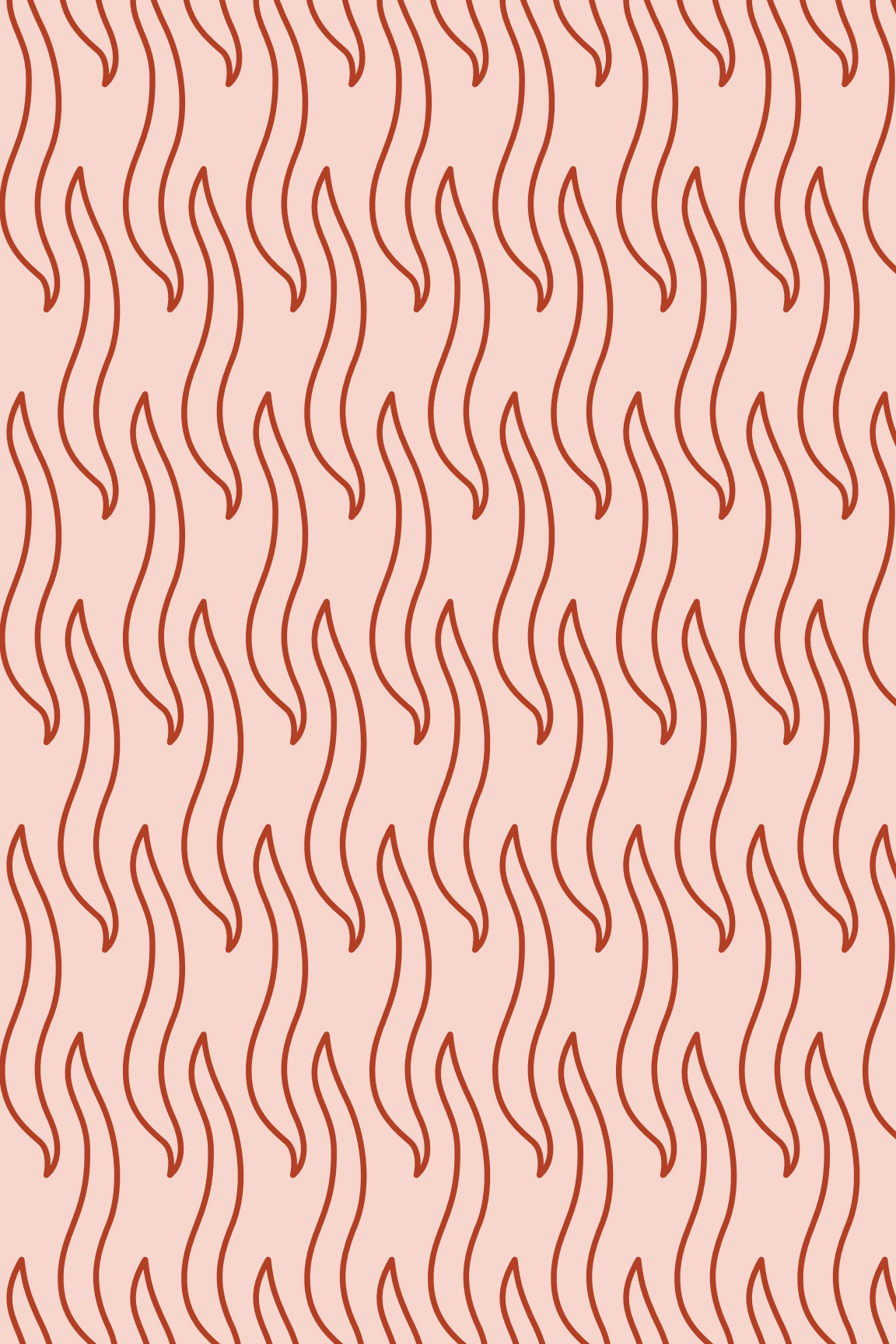
87. Ramsey SD, Willke RJ, Glick H, Reed SD, Augustovski F, Jonsson B, Briggs A, Sullivan SD. Cost-effectiveness analysis alongside clinical trials II-An ISPOR Good Research Practices Task Force report. *Value Health*. 2015;18(2):161-72.

88. Adetugbo K and Williams H. How well are randomized controlled trials reported in the dermatology literature? *Arch Dermatol*. 2000;136(3):381-5.

89. Montané E, Vallano A, Vidal X, Aguilera C, Laporte JR. Reporting randomised clinical trials of analgesics after traumatic or orthopaedic surgery is inadequate: a systematic review. *BMC Clin Pharmacol*. 2010;10:2. Published 2010 Jan 12.

90. van Dongen JM, El Alili M, Varga AN, Guevara Morel AE, Jornada Ben A, Khorrami M, van Tulder MW, Bosmans JE. What do national pharmacoeconomic guidelines recommend regarding the statistical analysis of trial-based economic evaluations? *Expert Rev Pharmacoecon Outcomes Res*. 2020;20(1):27-37.

91. Ramsey S, Willke R, Briggs A, Brown R, Buxton M, Chawla A, Cook J, Glick H, Liljas B, Petitti D, Reed S. Good research practices for cost-effectiveness analysis alongside clinical trials: the ISPOR RCT-CEA Task Force report. *Value Health*. 2005;8(5):521-33.



A

Relationship between low back pain, outcomes, and costs

2

The longitudinal relationships between pain severity and disability versus health-related quality of life and costs among chronic low back pain patients

Mutubuki, E.N.,[1] Beljon, Y., [1] Maas, E.T.,[1,2] Huygen, F.J.P.M., [3] Ostelo, R.W.J.G.,[1,4] van Tulder, M.W.,[1,5] van Dongen, J.M.[1]

[1] Department of Health Sciences, Faculty of Science, Vrije University Amsterdam, Amsterdam Movement Sciences, the Netherlands
[2] School of Population and Public Health, University of British Columbia, Canada
[3] Department of Anesthesiology, Centre of Pain Medicine, ErasmusMC, Rotterdam,
[4] Department of Epidemiology and Biostatistics, VU University Medical Center, Amsterdam Movement Sciences, the Netherlands
[5] Department of Physiotherapy & Occupational Therapy, Aarhus University Hospital, Aarhus, Denmark

Mutubuki EN, Beljon Y, Maas ET, et al. The longitudinal relationships between pain severity and disability versus health-related quality of life and costs among chronic low back pain patients. Qual Life Res. 2020;29(1):275-287. doi:10.1007/s11136-019-02302-w

ABSTRACT

Purpose: Previous studies found higher levels of pain severity and disability to be associated with higher costs and lower health-related quality of life. However, these findings were based on cross-sectional data and little is known about the longitudinal relationships between pain severity and disability versus health-related quality of life and costs among chronic low back pain patients. This study aims to cover this knowledge gap by exploring these longitudinal relationships in a consecutive cohort.

Methods: Data of 6,316 chronic low back pain patients were used. Measurements took place at 3, 6, 9, and 12 months. Pain severity (Numeric Pain Rating Scale; Range:0-100), disability (Oswestry Disability Index; Range:0-100), health-related quality of life (EQ-5D-3L: Range:0-1), societal and healthcare costs (cost questionnaire) were measured. Using linear generalized estimating equation analyses, longitudinal relationships were explored between: 1) pain severity and health-related quality of life, 2) disability and health-related quality of life, 3) pain severity and societal costs, 4) disability and societal costs 5) pain severity and healthcare costs, and 6) disability and healthcare costs.

Results: Higher pain and disability levels were statistically significantly related with poorer health-related quality of life (pain intensity:-0.0041; 95%CI:-0.0043 to -0.0039; disability:-0.0096; 95%CI:-0.0099 to -0.0093), higher societal costs (pain intensity:5; 95%CI:4 to 6; disability:17; 95%CI:14 to 20) and higher healthcare costs (pain intensity:2; 95%CI:2 to 3; disability:8; 95%CI:6 to 9).

Conclusion: Pain and disability were longitudinally related to health-related quality of life, societal costs, and healthcare costs. Disability had a stronger association with all outcomes compared to pain.

Keywords: pain, disability, health-related quality of life, societal costs, longitudinal analysis, low back pain

INTRODUCTION

Low back pain (LBP) is a highly prevalent health complaint. In 2015, the global point prevalence of activity-limiting LBP was estimated at 7.3%, implying that about 540 million people worldwide were affected by LBP at that moment in time [1]. Previous studies reported the lifetime-prevalence of LBP to range from 60 to 85% [2-5]. This indicates that people have a high probability of developing an LBP episode at any time during their life. In the upcoming decades, the aging of the population will likely lead to an increased prevalence of LBP as well as an increased number of patients whose pain persists for a period longer than 3 months (also defined as chronic LBP) [6,7].

Chronic LBP is associated with high pain levels, significant physical limitations, poorer prognosis, lower health-related quality of life and disability [3,8-10]. Around 57 million years lived with disability were found to be associated with LBP in 2016, and these have increased by more than 50% since 1990 [11]. Chronic LBP patients report quality of life scores that are comparable to those individuals with a life-threatening diagnosis [12]. Even though only 10-15% of LBP patients develop chronic LBP, research suggests that chronic LBP is responsible for the majority of LBP-related societal costs [6]. In the Netherlands, these LBP-related societal costs were estimated to be as high as 3.5 billion euros in 2007, which equals about 0.6% of the Dutch gross national product (GNP) [6]. In the United States, the estimated annual total societal cost of LBP was estimated at 100 billion dollars [13,14]. Absenteeism, early retirement, and a loss of productivity while being at work are the most important drivers of these societal costs [15].

Previous studies found a higher level of pain severity and/or disability to be related to higher costs and a lower health-related quality of life [16,10,17-19]. A study by Horng et al., for example, reported significant correlations between pain intensity and disability and health-related quality of life [17,18]. Long lasting, persisting pain and functional limitations that LBP patients experience can cause disability and interfere with their quality of life [17,20]. Chiarotto et al., reported a positive correlation between pain severity, as measured using a Numeric Rating

Scale (NRS), and disability and a negative correlation between pain severity, as measured using the Brief Pain Inventory-Pain Severity, and health-related quality of life [21]. Sadosky et al. found that an increasing pain severity level was associated with higher indirect costs (i.e. productivity-related costs), direct costs (i.e. healthcare costs), and societal costs amongst Japanese LBP patients [19].

Previous studies on the relation between pain severity and disability versus health-related quality of life and healthcare and societal costs among chronic LBP patients were cross-sectional in nature [19]. This means that they explored whether pain severity and/or disability were associated at a certain point in time with health-related quality of life and/or healthcare and societal costs. Such cross-sectional studies do not provide insight into whether individual changes in one variable (e.g. pain severity) are related to individual changes in another (e.g. costs). Such relationships can only be studied using a longitudinal study design, in which both variables are measured and compared over time [22].

This study aims to cover this knowledge gap by exploring the longitudinal relationships between pain severity and disability versus health-related quality of life, healthcare and societal costs among chronic LBP patients. Based on previous cross-sectional research, we expect that higher pain and disability are associated with reduced health-related quality of life (negative longitudinal relationship) and higher healthcare and societal costs (positive longitudinal relationship). Next to providing valuable information for clinical practice, information on the longitudinal relationships between pain severity and disability versus health-related quality of life and costs amongst chronic LBP patients could provide valuable input for health economic modelling studies in the area of chronic LBP.

METHODS

Study population and design

Data collected during the MinT (minimal invasive treatment) study [23] were used to explore the longitudinal relationships between pain severity and disability versus health-related quality of life and costs among chronic LBP patients. The MinT study was conducted in the Netherlands, and consisted of four randomized controlled trials and an observational study. The overall aim of the MinT study was to assess the effectiveness and cost-effectiveness of adding minimal interventional procedures to a standardized treatment program, compared with a standardized treatment program alone [23,24]. A detailed description of the MinT study can be found elsewhere [23]. In the present study, only data of chronic LBP patients participating in the observational branch of the MinT study were used (i.e. patients experiencing LBP symptoms for more than 12 weeks). In order to be eligible to participate in the observational study, and thus to be included in the present study, patients had to be aged between 18 and 70 years, referred to a pain clinic with suspected chronic mechanical LBP and without improvement of symptoms after conservative treatment [23]. The observational study monitored patients who did not want to, or were not eligible, to participate in the aforementioned randomized controlled trials [23].

Outcome Measures

Dependent variables: Health-related quality of life, societal costs, and healthcare costs

Three dependent variables were used in this study, all of which were measured at 3, 6, 9, and 12-month follow-up. Health-related quality of life was also measured at baseline, whereas healthcare and societal costs were not. To improve comparability across the analyses, only follow-up measurement values were used for assessing the longitudinal relationships.

1) *Health-related quality of life:* Health-related quality of life was measured using the EQ-5D-3L. The EQ-5D-3L is a health-

related quality of life scale that has previously been found to be responsive amongst chronic LBP patients [25]. The EQ-5D-3L consists of five dimensions of health, including mobility, self-care, daily activities, pain/discomfort, and anxiety/depression, each with three levels of severity. The participants' EQ-5D-3L scores were converted into utility values using the Dutch tariff [26]. Utility values are preference weights, indicating a person's value or desirability of a certain health state on a scale anchored at 0 (equal to death) and 1 (equal to full health) [27].

2) *Societal costs:* Comprised in societal costs were healthcare, informal care, unpaid productivity and work absenteeism costs. Resource use was measured using cost questionnaires [28]. Healthcare use included the use of primary care (e.g. visits to a general practitioner or physiotherapist) and secondary care (e.g. visits to a medical specialist or pain clinic). Data from the updated Dutch Manual of Costing were used to value costs of common healthcare interventions, such as appointments with a general physician and a physical therapist [29]. Costs of less common interventions were estimated using an average of five quotes from various practitioners across the country and/or pricelists of professional organisations. Informal care and unpaid productivity were valued using a recommended Dutch shadow price [29]. To measure work absenteeism, the Productivity and disease Questionnaire (PRODISQ) was used [30]. Absenteeism costs were estimated in accordance with the friction cost approach and using gender-specific price weights provided by the updated Dutch Manual of Costing [29]. All cost categories were measured with 3-month recall periods [31].
3) *Healthcare costs:* Comprised in healthcare costs were primary and secondary healthcare costs. The measurement and valuation of healthcare costs has been outline above.

Independent variables: Pain severity and disability

Two independent variables were used in this study, both of which were measured at baseline, 3, 6, 9, and 12-month follow-up:

1) *Pain intensity:* Pain severity was measured using the NPRS (range 0 - no pain to 10 - worst pain imaginable). Scores were

transformed to a 0-100 scale to improve the interpretation and comparability of outcomes. Several studies concluded that the validity and sensitivity of the NPRS was appropriate for measuring pain in chronic LBP patients [32,33]. A clinically meaningful change for people with LBP on the NPRS was previously found to be two (equalling 20 on the 0-100 scale) [34].

2) *Disability:* Disability was measured using the Oswestry Disability Index (ODI: range 0 - no disability to 100-maximum disability possible). The ODI is a commonly used outcome measure amongst LBP patients [35-38] and is reported to be a valid, reliable and responsive hence suitable as a clinical measure [35]. A clinically meaningful change for people with LBP on the ODI was previously found to be ten points on the 0-100 point ODI [39].

The ODI and NPRS are both part of the core outcome set recommended for LBP [40].

Potential confounding factors

Potential confounding factors included were based on literature [41] and measured at baseline. These included:

- Patient expectations (Credibility/Expectancy Questionnaire [CEQ] [42]; range 0-least credibility/expectancy to 100 – more credibility/expectancy).
- Pain severity (Numeric Pain Rating Scale [NPRS]; range 0 - no pain to 100 - worst pain imaginable) [34]. For the purpose of this study, scores were transformed to 0-100. (In the analyses in which disability and health-related quality of life were included)
- Disability (Oswestry Disability Index [ODI]; range 0 - no disability to 100 – maximum disability) [43,37]. (In the analyses in which pain and health-related quality of life were included)
- Health-related quality of life (EuroQol [EQ-5D-3L]; range 0-equal to death 1 – equal to full health) [44].
- General health – mental component score and physical component score (Rand-36 [Rand-36]; scores range 0- lowest general health to 100- highest general health) [45-47]. The two component scores were assessed for being a confounding

variable separately.
- Impact of pain experience (Multidimensional Pain Inventory [MPI]; range 0- least/best to 100 most/worst) [48,49].
- Education level (low/moderate/high). Low-indicates, no education, primary level education, lower vocational and lower secondary education, moderate-indicates higher secondary education or undergraduate, high-indicates tertiary education university or postgraduate).
- Body Mass Index ([BMI], weight in kg/(height in meters)2).
- Employment (yes/no).
- Recurrent complaints (yes/no).
- Age (years).
- Gender (male/female).
- Nationality (Dutch/non-Dutch).
- Smoking (yes/no).
- Type of health care insurance (basic/additional).
- Region of residence (south/north/ east/west).
- Married/living together yes/no).
- Diagnosis (sacroiliac joint (SI)/facet/disc/ combined/ unclear).

Statistical analysis

The patients' baseline characteristics were descriptively summarized. Missing data were handled using multiple imputation to avoid possible bias due to selective drop-out of participants [50]. Imputations were performed using the Multiple Imputation by Chained Equations algorithm with predictive mean matching [51]. The imputation model included all available potential confounders, pain intensity, disability, health-related quality of life, and cost values.

For answering the research question, linear generalized estimating equation (GEE) analyses were performed. A GEE analysis is a so-called sophisticated longitudinal data analysis technique, in which the relationship between the variables in the model (e.g. pain severity and societal costs) at different time points (i.e. 3, 6, 9, and 12 months) is analyzed simultaneously. Herewith, the estimated regression coefficient reflects the longitudinal relationship between the dependent variable (e.g. societal costs) and the independent variable(s) (e.g.

pain severity), using all available data, and thus providing an indication of whether changes in the dependent variable are related to changes in the independent variable [22] within and between participants over different measurement time points. Six separate longitudinal relationships were assessed between: 1) pain severity and health-related quality of life, 2) disability and health-related quality of life, 3) pain severity and societal costs, 4) disability and societal costs 5) pain severity and healthcare costs, and 6) disability and healthcare costs. Longitudinal relationships 1) and 2) were explored with a Gaussian distribution and an identity link. Longitudinal relationships 3) to 6) were explored with a gamma distribution and an identity link. The gamma distribution was chosen to take into account the right skewed nature of cost data. In all of the analyses, an exchangeable correlation structure was assumed. First, crude analyses were performed that solely included the dependent and the independent variables. Second, adjusted analyses were performed that also included potential confounding factors. Variables that changed the regression coefficient by more than 10% were deemed confounders and were included in the model. All analyses were performed in Stata (version 14 SE, Stata Corp). Statistical significance was set on $p<0.05$.

RESULTS

Participants

Data from 6,316 chronic LBP patients were analyzed in the present study. Of them, the majority were female (66%), overweight (67%), Dutch (95%), had a low level of education (56%), had a mean age of 57 years and more than half were unemployed (59%) (Table 1). Cost data had the highest percentage of missing data and most data was missing at 9-month follow-up. A detailed description of the percentages of missing data per outcome and per time point can be found in Figure 1.

Table 1: Overview of the data collection

Participant characteristic	All patients
Age (years) [mean (SD)]	57.2 (13,4)
Gender [n (%)]	
Female	4142 (66)
Male	2093 (34)
BMI [n (%)]	
BMI<18.5 (underweight)	37 (1)
BMI≥18.5<25 (normal weight)	1687 (32)
BMI ≤25<30 (overweight)	2060 (39)
BMI≥30 (obese)	1463 (28)
Smoking [n (%)]	
Yes	1413 (26)
No	3920 (73)
Educational level [n (%)]	
Low (no education, primary level education, lower vocational and lower secondary education)	2925 (56)
Moderate (higher secondary education or undergraduate)	1467 (28)
High (tertiary, university level, postgraduate)	830 (16)
Living together with a partner [n (%)]	
Yes	4663 (75)
No	1593 (26)
Nationality [n (%)]	
Dutch	5049 (95)
Non-Dutch:	278(5.2)
Surinamese	*21 (0.4)*
Antillean/Aruban	*22 (0.4)*
Turkish	*63 (1)*
Moroccan	*42 (1)*
Other	*130 (2.4)*
Region in the Netherlands [n (%)]	
South	2029 (32)
North	1165 (19)
East	1280 (20)
West	1782 (28)
Employment [n (%)]	
Yes	1687(42)
No	2376 (59)

Continued - Table 1: Overview of the data collection

Recurrent low back pain [n (%)]	
Yes	3,174 (63)
No	1876 (37)
Diagnosis-source of pain [n (%)]	
1= SI	1864 (33)
2= Facet	2269 (41)
3=Disc	18 (0.3)
4= Combined	1391 (25)
5= Unclear	66 (1)
Patients expectations	
Credibility [mean (SD)] range 0-100	77.1 (17.5)
Expectancy [mean (SD)] range 0-100	57.8 (17.3)
Rand-36	
Mental [mean (SD)] range 0-100	22.6 (5)
Physical [mean (SD)] range 0-100	18.5 (4)
Health-related quality of life(utility) [mean (SD)] range 0-100	48 (29)
MPI [mean (SD)] range per subscale 0-100	
Pain severity	22.6 (5.7)
Interference with daily activities	5.8 (1.9)
Life control	21.2 (6.3)
Affective distress	15.4 (4.6)
Support	28.6 (7.6)
Type of health care insurance [n (%)]	
Basic insurance	633 (12)
Comprehensive (basic+additional cover)	4630 (86)
I don't know	55 (1)
ODI functional disability [mean (SD)] range 0-100	11.1 (9)
Pain severity[mean (SD)] range 0-100	73 (16)

Note: percentages have been rounded off hence values a bit less than 100% and a bit more that 100%
Scores for MPI, Rand 36, patient expectations, health related quality of life were transformed to a range of 0-100 to enable comparability.
Diagnosis was based on patient history and physical examination
ODI-Oswestry Disability Index
MPI-Multidimensional Pain Inventory

Disability

Table 2 shows the results from the longitudinal analyses between disability and health-related quality of life, societal costs and healthcare costs. Disability and health-related quality of life had a statistically significant negative longitudinal relationship (B: -0.0096; 95%CI: -0.0099 to -0.0093). As none of the possible confounding factors changed the regression coefficient by more than 10% an adjusted model was not required. The crude analysis using disability and societal costs suggested a significant positive longitudinal relationship (B: 24; 95%CI: 20 to 29). After adjusting for confounding, the identified longitudinal relationship between disability and societal costs remained statistically significant (B: 17; 95%CI: 14 to 20). A significant positive longitudinal relationship was also observed in both the crude (B: 10; 95%CI: 9 to 12) and adjusted (B: 8; 95%CI: 6 to 9) analyses using disability and healthcare costs.

Pain

Table 3 shows the results from the longitudinal analyses between pain and health-related quality of life, societal costs and healthcare costs. A significant negative longitudinal relationship was observed between pain and health-related quality of life. After adjusting for confounding, the results still suggested a significant negative longitudinal relationship between pain and healthcare costs (B: -0.0041; 95%CI: -0.0043 to -0.0039). The crude analyses using pain and societal costs suggested a significant positive longitudinal relationship (B: 8; 95% CI; 6 to 10). After adjusting for confounding, the adjusted analysis also suggested a significant positive longitudinal relationship between pain and societal costs (B: 5; 95%CI: 4 to 6). Pain and healthcare costs also had a significant positive longitudinal relationship in both the crude (B: 3; 95%CI: 2 to 5) and adjusted analysis (B: 2; 95% CI: 2 to 3).

Table 2: Longitudinal analyses between disability, societal costs, healthcare costs and health-related quality of life

Results Disability	**Crude**			**Adjusted**		
		95% CI			95% CI	
Disability ODI (0-100)	Beta	Lower bound	Upper bound	Beta	Lower bound	Upper bound
*Total costs-societal perspective	24	20	29	17	14	20
**Total costs-health care perspective	10	9	12	8	6	9
***Health-related quality of life	-0.0096	-0.0099	-0.0093			

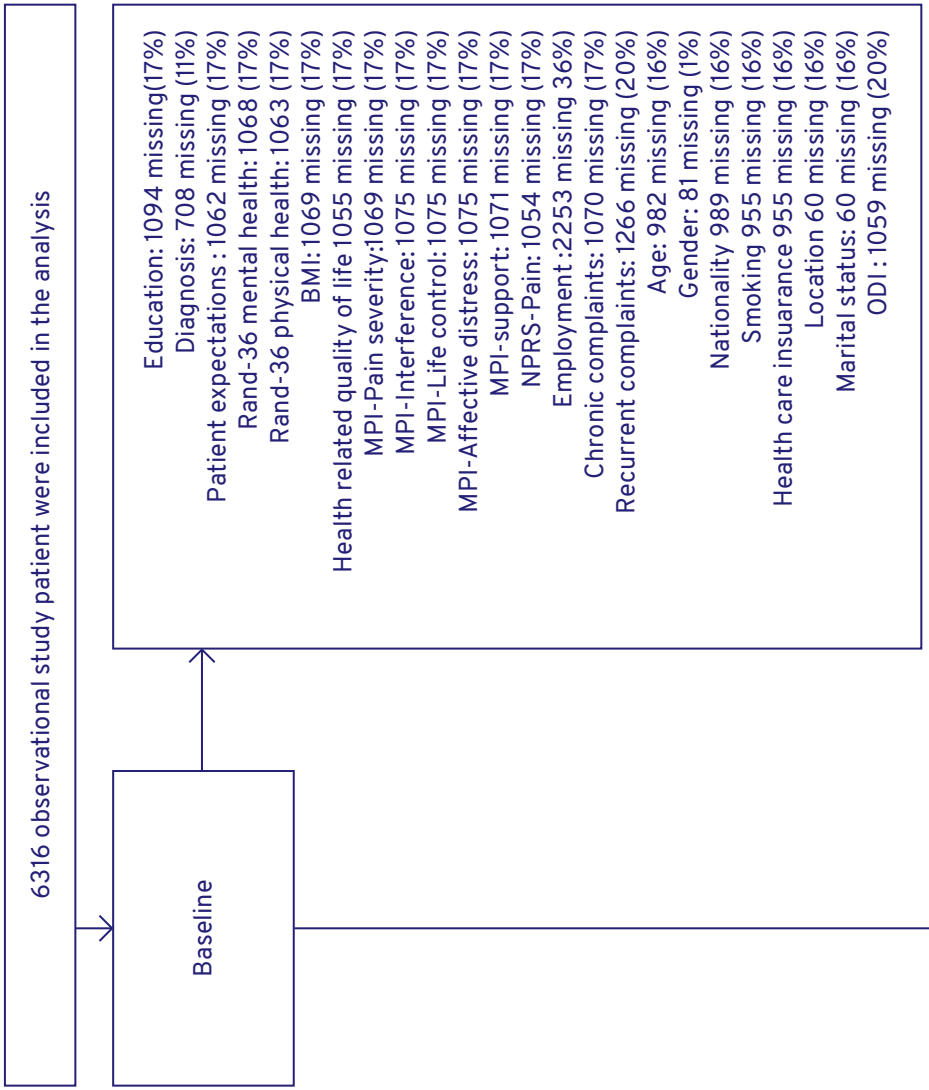
** Adjusted for health-related quality of life, physical health, MPI life control, MPI_Inteference, MPI_Pain severity, mental health, disability*

***Adjusted for physical health (SF-36), pain impact experience (MPI interference), health-related quality of life (EQ-5D)*

**** No confounding factors; none of the confounders changed the regression co-efficient by more than 10%*

Scores for health-related quality of life were transformed to a range of 0-100 to enable comparability

Figure 1: Flowchart of missing data at each follow-up moment



Continued - Figure 1: Flowchart of missing data at each follow-up moment

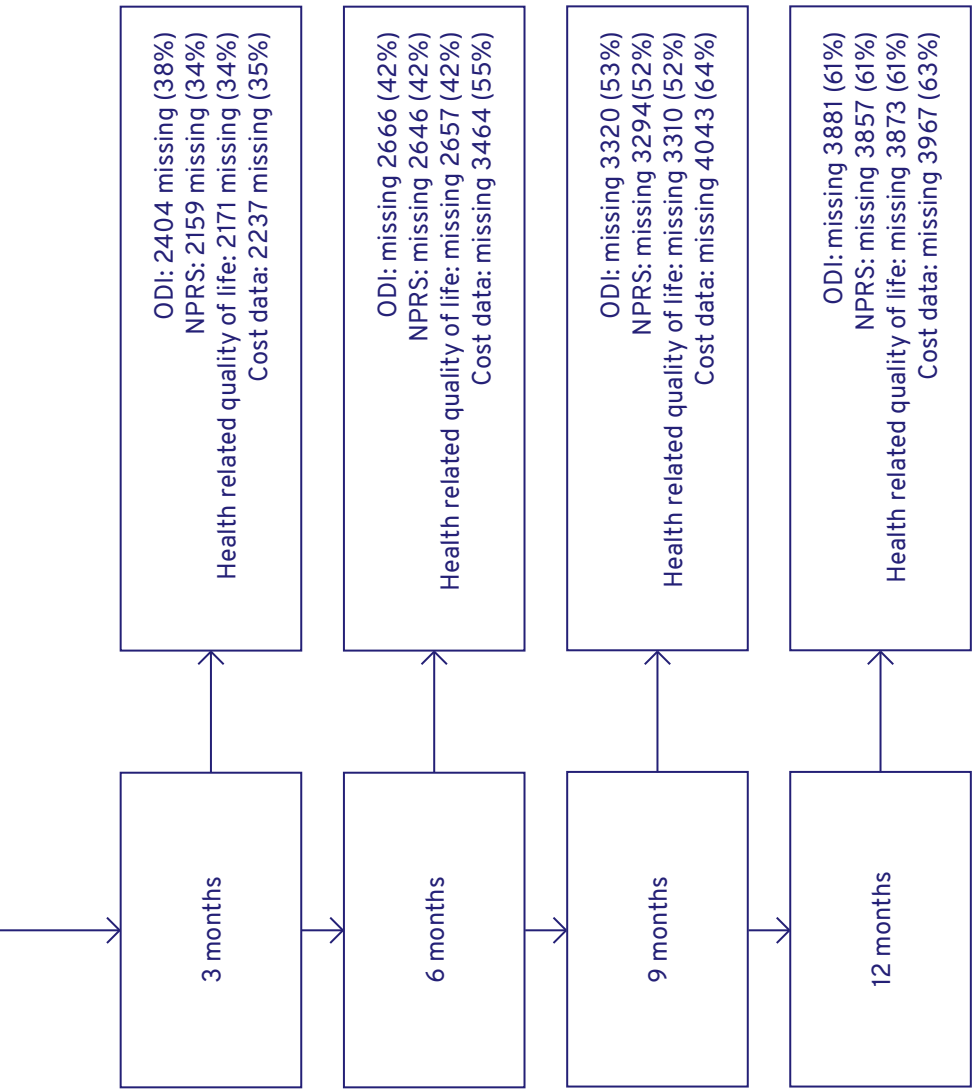


Table 3: Longitudinal analyses between pain, societal costs, healthcare costs and health-related quality of life

Results pain	**Crude**			**Adjusted**		
		95% CI			95% CI	
Pain (0-100)	Beta	Lower bound	Upper bound	Beta	Lower bound	Upper bound
*Total costs-societal perspective	8	6	10	5	4	6
**Total costs-health care perspective	3	2	5	2	2	3
***Health-related quality of life	-0.0041	-0.0043	-0.0038			

**Adjusted for health-related quality of life, physical health, MPI life control, mental health, disability*

*** Adjusted for MPI life control, MPI_Inteference, MPI_Pain severity, health-related quality of life, physical health (SF-36), mental health*

**** No confounding factors; none of the confounders changed the regression co-efficient by more than 10%*

DISCUSSION

Main findings

This study found pain severity and disability both to have a statistically significant negative longitudinal relationship with health-related quality of life, and a statistically significant positive longitudinal relationship with societal as well as healthcare costs. In GEE, regression coefficients have a double interpretation resulting in a pooled coefficient of a within-subject and a between-subject effect [22]. Interpreting these regression coefficients in terms of practical relevance indicates that a 1-point increase in disability, for example, is related to a 0.0096 point decrease in health-related quality of life (range: 0-1), 17 euros increase in societal costs and 8 euros increase in healthcare costs per three months. A clinically relevant increase in disability (defined as a 10 point increase on the 0-100 point ODI) [39] is thus associated with a decrease in health-related quality of life by 0.096 points (range: 0-1), and an increase in societal as well healthcare costs by 170 and 80 euros per 3-month period. Thus, the potential costs savings associated with relevant improvements in pain and disability are tremendous in a highly prevalent disorder such as low back pain. Moreover, a clinically relevant increase in disability was found to be longitudinally related to a more than clinically relevant increase in health-related quality of life, which was previously found to be equal to an increase of 0.057 or more [52,53]. For pain intensity, the associated decrease in health-related quality of life was slightly smaller than the established minimal clinically relevant difference for health-related quality of life (i.e. 0.0041 versus 0.059).

All of our findings were in line with our expectation that pain and disability would have a statistically significant negative relationship with health-related quality of life and a statistically significant positive longitudinal relationship with societal costs and healthcare costs. Also, it is noteworthy that the impact of pain on health-related quality of life and costs was found to be about 2.5 times smaller than the impact of disability on health-related quality of life and costs even though patients had high baseline scores of pain and relatively 'lower scores' on disability. This might suggest that it is not the level of pain severity that has

a strong association with an individual's health-related quality of life and/or costs but the way in which an individual's pain influences his or her daily activities. However, further research is needed to confirm this.

Comparison with literature

To the best of our knowledge, no studies have explored the longitudinal relationships between pain severity and disability versus health-related quality of life and costs. Nonetheless, a cross-sectional study by Sadosky et al. found an increasing pain severity to be related with a worsening health-related quality of life as well as increased healthcare and societal costs. This is in line with the findings of the present study. In contract to the present study, however, Sadosky et al. also included presenteeism costs (i.e. costs related to reduced productivity while being at work) and their study was conducted among acute as well as chronic LBP patients instead of chronic LBP patients only. Like the present study, a study of Stefane et al. found a significant negative association between pain and disability versus health-related quality of life. In line with the current study, Stefane et al. found health-related quality of life to be more strongly associated to disability than to pain. Unlike the present longitudinal study, the study of Stefane et al. was cross-sectional in nature and disability was measured using the Ronald- Morris 24 items questionnaire [54] instead of using the ODI.

In our study, disability was found to have about 2.5 times higher impact on health-related quality of life, societal costs, and healthcare costs. Our reasoning that, this might suggest that, it is not the level of pain severity that has a strong association with an individual's health-related quality of life and/or costs but the way in which an individual's pain influences his or her daily activities is supported by a study of Horng et al. In their study, Horng et al., reported that pain persistence and limitation of activities for daily living had more influence on a patient's health-related quality of life compared to pain severity alone in both acute and chronic LBP patients [17]. Our reasoning is further supported by Lame et al. who reported pain catastrophizing as the most important predictor of individual health-related quality of life in a heterogeneous group of chronic pain patients. In

their study pain catastrophizing had the strongest association with individuals health-related quality of life compared to pain severity and chronic LBP patients had the lowest quality of life [55]. Pain catastrophizing is generally defined as excessive negative orientation towards pain/ noxious stimuli [56,57]. High levels of pain catastrophizing were associated with disability, poor outcomes and pain severity for patients with LBP [57-59].

Strengths and Limitations

Strengths of the present study include that it is the first study to use a longitudinal design to explore whether relationships exist between pain severity and disability versus health-related quality of life and costs. In addition, the large cohort of observed patients with chronic LBP patients (n=6,316) greatly increases the power of this study. Another advantage is the use of imputation methods to deal with missing data thereby avoiding complete-case analysis, which would have significantly reduced the study's power and precision. Multiple imputation is the preferred statistical method for dealing with missing, particularly when costs are involved [50].

Limitations of the present study include the absence of presenteeism costs in the analyses, whereas presenteeism more than absenteeism is reported to be disproportionately affected by pain [16]. As the results of Sadosky et al., who did include presenteeism costs, were in line with those of the present study, we do not expect the absence of presenteeism costs to have greatly biased our conclusion. Nonetheless, future studies should include presenteeism costs to give a more accurate representation of true costs related to lost productivity. Second, there is an over representation of females (66.4%) in the present study, in contrast with the percentage of women with LBP in the Netherlands (56%). This could have resulted in an underestimation of costs since men earn more than women [60] and tend to use more healthcare for LBP [61]. Nonetheless, as stratified post-hoc analyses indicated that, except for one unadjusted analyses, all longitudinal relationships were statistically significant amongst men and women with similar beta coefficients (Appendix 1) , we do not expect the overrepresentation of women to have severely biased our

results and conclusions. Future studies should include a larger representation of males, reflecting the 44% of males suffering from chronic LBP, to enable better generalizability of our results. Third, although mainly valid and reliable questionnaires were used, the self-reported nature of the questionnaires might have caused recall and or social desirability bias. We tried limiting the recall bias by minimizing the recall period to three months [29]. As it seems unlikely that recall-bias or the degree to which participants gave socially desirable answers systematically differed over time, it is not expected that self-report biased the results. Fourth, lack of comorbidity factors, which could have been potential confounding factors could have led to underestimation of costs and the impact on health-related quality of life, since confounding could not controlled for. Fifth, in the present study, the EQ-5D-3L was used to measure health-related quality of life, whereas since the inception of the MinT study [23], an updated five level version of the EQ-5D has been published [62]. However, as both have previously found to be valid means to measure health-related quality of life, we do expect our reliance on the EQ-5D-3L to have biased our results [63]. Also, even though GEE analysis offers an efficient means to analyse the longitudinal relationship between variables, its results may heavily depend on the assumptions made. That is, with GEE analysis, the adjustment for time is carried out by assuming a priori a certain "working" correlation structure for the repeated measurements. Even though GEE analysis is assumed to be robust against a wrong choice of correlation structure, evidence suggests that results may differ extensively across correlation structures [22]. Based on the recommendations of Twisk et al., we assumed an "exchangeable" correlation structure, in which correlations between subsequent measurements are assumed to be equal irrespective of the length of the time intervals [22]. To assess the robustness of the current findings to the choice of correlation structure, we performed a post-hoc analysis with an "unstructured" correlation structure, in which no particular structure is assumed and all possible correlations between repeated measurements have to be estimated [22]. As the results of the post-hoc analysis are in line with those of the main analysis (Appendix 1), we consider the current findings to be robust against the choice of correlation structure.

Implications for practice and research

Our findings indicate that the potential costs savings associated with relevant improvements in pain and disability are tremendous in a prevalent disorder such as LBP. A clinical improvement in disability, 10 points on the 0-100 point ODI [39], will result in potential savings of 170 per LBP patient per 3 months. Our study also provides some preliminary evidence, that is, disability is more associated with higher societal and healthcare costs and poorer health-related quality of life, than pain severity. Further research into this topic is warranted, but for now these findings at least suggest that focussing initiatives and interventions on disability more that pain severity may improve patient outcomes, i.e. health-related quality of life and costs. Also, the aim of the present study was to explore the separate relationships of pain and disability with healthcare costs, societal costs and health-related quality of life. Therefore, the combined influence of pain severity and disability on costs and health-related quality of life was not explored, neither were the potential interactions between pain and disability. This should be explored further in future research.

CONCLUSION

The present study showed that both pain severity and disability are longitudinally related to health-related quality of life, societal costs, and healthcare costs. Disability had a stronger association with all outcomes compared to pain suggesting that, it is not the level of pain severity that influences the height of an individual's health-related quality of life and costs, but the way in which an individual's pain influences his or her daily activities.

Funding
This study was funded by grant 171202013 from the Netherlands Organization for Health Research and Development, by the Dutch Society for Anesthesiology, and the Dutch health insurance companies.

Conflict of interest

EM wrote the initial version of the manuscript. EM, JB, JvD, and ET were involved in the data analysis process. All authors reviewed and commented on the manuscript. FH, MvT, and RO received funding for the study. The authors have no conflicts of interest to declare.

Ethical approval

"All procedures performed in studies involving human participants were in accordance with the ethical standards of the Medical Ethics Committee of the Erasmus Medical Centre in Rotterdam registration number MEC-2012-079 and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards." Local research governance was obtained from all participating pain clinics.

Informed consent

"Informed consent was obtained from all individual participants included in the study."

Appendix 1a: Longitudinal analyses between disability, societal costs, healthcare costs and health-related quality of life among men

Results Disability	**Crude**			**Adjusted**		
		95% CI			95% CI	
Disability ODI (0-100)	Beta	Lower bound	Upper bound	Beta	Lower bound	Upper bound
*Total costs-societal perspective	23	20	25	15	12	18
**Total costs-health care perspective	9	7	11	5	4	7
***Health-related quality of life	-0.0094	-0.0098	-0.0091			

** Adjusted for health-related quality of life, physical health, MPI life control, MPI_Inteference, MPI_Pain severity, mental health, disability*
***Adjusted for physical health (SF-36), pain impact experience (MPI interference), health-related quality of life (EQ-5D)*
**** No confounding factors; none of the confounders changed the regression co-efficient by more than 10%*
Scores for health-related quality of life were transformed to a range of 0-100 to enable comparability

Appendix 1b: Longitudinal analyses between pain, societal costs, healthcare costs and health-related quality of life among men

Results Pain	**Crude**			**Adjusted**		
		95% CI			95% CI	
Pain (0-100)	Beta	Lower bound	Upper bound	Beta	Lower bound	Upper bound
*Total costs-societal perspective	4	2	7	2	1	4
**Total costs-health care perspective	2	-0.4	3	1	0.2	2
***Health-related quality of life	-0.0039	-0.0041	-0.0037			

**Adjusted for health-related quality of life, physical health, MPI life control, mental health, disability*
*** Adjusted for MPI life control, MPI_Inteference, MPI_Pain severity, health-related quality of life, physical health (SF-36), mental health*
**** No confounding factors; none of the confounders changed the regression co-efficient by more than 10%*

Appendix 1c: Longitudinal analyses between disability, societal costs, healthcare costs and health-related quality of life among women

Results Disability	**Crude**			**Adjusted**		
		95% CI			95% CI	
Disability ODI (0-100)	Beta	Lower bound	Upper bound	Beta	Lower bound	Upper bound
*Total costs-societal perspective	25	21	30	18	15	22
**Total costs-health care perspective	11	9	13	9	7	10
***Health-related quality of life	-0.0096	-0.0099	-0.0094			

** Adjusted for health-related quality of life, physical health, MPI life control, MPI_Inteference, MPI_Pain severity, mental health, disability*
***Adjusted for physical health (SF-36), pain impact experience (MPI interference), health-related quality of life (EQ-5D)*
**** No confounding factors; none of the confounders changed the regression co-efficient by more than 10%*
Scores for health-related quality of life were transformed to a range of 0-100 to enable comparability

Appendix 1d: Longitudinal analyses between pain, societal costs, healthcare costs and health-related quality of life among women

Results Pain	**Crude**			**Adjusted**		
		95% CI			95% CI	
Pain (0-100)	Beta	Lower bound	Upper bound	Beta	Lower bound	Upper bound
*Total costs-societal perspective	8	7	10	6	5	7
**Total costs-health care perspective	4	3	4	2	2	3
***Health-related quality of life	-0.0039	-0.0041	-0.0038			

**Adjusted for health-related quality of life, physical health, MPI life control, mental health, disability*
*** Adjusted for MPI life control, MPI_Inteference, MPI_Pain severity, health-related quality of life, physical health (SF-36), mental health*
**** No confounding factors; none of the confounders changed the regression co-efficient by more than 10%*

Appendix 2a: Longitudinal analyses between disability, societal costs, healthcare costs and health-related quality of life – using a unstructured correlation structure

Results Disability	**Crude**			**Adjusted**		
		95% CI			95% CI	
Disability ODI (0-100)	Beta	Lower bound	Upper bound	Beta	Lower bound	Upper bound
*Total costs-societal perspective	25	21	28	17	14	21
**Total costs-health care perspective	11	9	12	8	6	9
***Health-related quality of life	-0.0096	-0.0099	-0.0093			

** Adjusted for health-related quality of life, physical health, MPI life control, MPI_Inteference, MPI_Pain severity, mental health, disability*
***Adjusted for physical health (SF-36), pain impact experience (MPI interference), health-related quality of life (EQ-5D)*
**** No confounding factors; none of the confounders changed the regression co-efficient by more than 10%*
Scores for health-related quality of life were transformed to a range of 0-100 to enable comparability

Appendix 2b: Longitudinal analyses between pain, societal costs, healthcare costs and health-related quality of life – using an unstructured correlation structure

Results Pain	**Crude**			**Adjusted**		
		95% CI			95% CI	
Pain (0-100)	Beta	Lower bound	Upper bound	Beta	Lower bound	Upper bound
*Total costs-societal perspective	8	7	10	5	4	7
**Total costs-health care perspective	4	3	5	2	2	3
***Health-related quality of life	-0.0041	-0.0043	-0.0039			

**Adjusted for health-related quality of life, physical health, MPI life control, mental health, disability*
*** Adjusted for MPI life control, MPI_Inteference, MPI_Pain severity, health-related quality of life, physical health (SF-36), mental health*
**** No confounding factors; none of the confounders changed the regression co-efficient by more than 10%*

REFERENCES

1. Hartvigsen, J., Hancock, M. J., Kongsted, A., Louw, Q., Ferreira, M. L., Genevay, S., et al. (2018). What low back pain is and why we need to pay attention. *The Lancet*.

2. Hoy, D., March, L., Brooks, P., Woolf, A., Blyth, F., Vos, T., et al. (2010). Measuring the global burden of low back pain. *Best Pract Res Clin Rheumatol*, 24(2), 155-165, doi:10.1016/j.berh.2009.11.002.

3. Andersson, G. B. (1999). Epidemiological features of chronic low-back pain. *Lancet, 354*(9178), 581-585, doi:10.1016/S0140-6736(99)01312-4.

4. Krismer, M., van Tulder, M., Low Back Pain Group of the, B., & Joint Health Strategies for Europe, P. (2007). Strategies for prevention and management of musculoskeletal conditions. Low back pain (non-specific). *Best Pract Res Clin Rheumatol*, 21(1), 77-91, doi:10.1016/j.berh.2006.08.004.

5. Manchikanti, L., Singh, V., Datta, S., Cohen, S. P., Hirsch, J. A., & American Society of Interventional Pain, P. (2009). Comprehensive review of epidemiology, scope, and impact of spinal pain. *Pain physician*, 12(4), E35-70.

6. Lambeek, L. C., van Tulder, M. W., Swinkels, I. C., Koppes, L. L., Anema, J. R., & van Mechelen, W. (2011). The trend in total cost of back pain in The Netherlands in the period 2002 to 2007. *Spine (Phila Pa 1976), 36*(13), 1050-1058, doi:10.1097/BRS.0b013e3181e70488.

7. Smith, M., Davis, M. A., Stano, M., & Whedon, J. M. (2013). Aging baby boomers and the rising cost of chronic back pain: secular trend analysis of longitudinal Medical Expenditures Panel Survey data for years 2000 to 2007. *J Manipulative Physiol Ther, 36*(1), 2-11, doi:10.1016/j.jmpt.2012.12.001.

8. Semeru, G. M., & Halim, M. S. (2019). Acceptance versus

catastrophizing in predicting quality of life in patients with chronic low back pain. *Korean J Pain, 32*(1), 22-29, doi:10.3344/kjp.2019.32.1.22.

9. Costa, L. D. C. M., Maher, C. G., McAuley, J. H., Hancock, M. J., Herbert, R. D., Refshauge, K. M., & Henschke, N. (2009). Prognosis for patients with chronic low back pain: inception cohort study. . *Bmj, 339*(b3829).

10. Vlaeyen J.W.S MCG, W. K., van Zundert J, Beraldo Meloto C, Diatchenko L, Michele C. Battié M.C, Goossens M, Koes B, Linton S. J. (2018). Low back pain *Nature Reviews Disease Primers 4*(52).

11. Vos, T., Abajobir, A. A., Abate, K. H., Abbafati, C., Abbas, K. M., Abd-Allah, F., ... & Aboyans, V. (2017). Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet, 390*(10100), 1211-1259.

12. Fredheim, O. M. S., Kaasa, S., Fayers, P., Saltnes, T., Jordhøy, M., & Borchgrevink, P. C. (2008). Chronic non- malignant pain patients report poor health related quality of life as palliative cancer patients. *Acta Anaesthesiologica Scandinavica, 52*(1), 143-148.

13. Dieleman, J. L., Baral, R., Birger, M., Bui, A. L., Bulchis, A., Chapin, A., et al. (2016). US spending on personal health care and public health, 1996-2013. *Jama, 316*(24), 2627-2646.

14. Katz, J. N. (2006). Lumbar disc disorders and low-back pain: socioeconomic factors and consequences. *Journal of bone and joint surgery, 88*, 21-24.

15. Tsuboi, Y., Murata, S., Naruse, F., & Ono, R. (2019). Association between pain-related fear and presenteeism among eldercare workers with low back pain. *Eur J Pain, 23*(3), 495-502.

16. Sadosky, A. B., Taylor-Stokes, G., Lobosco, S., Pike, J., & Ross, E. (2013). Relationship between self-reported low-back pain

severity and other patient-reported outcomes: results from an observational study. *Clinical Spine Surgery, 26*(1), 8-14.

17. Horng, Y. S., Hwang, Y. H., Wu, H. C., Liang, H. W., Mhe, Y. J., Twu, F. C., et al. (2005). Predicting health-related quality of life in patients with low back pain. *Spine (Phila Pa 1976), 30*(5), 551-555.

18. Von Korff, M., Crane, P., Lane, M., Miglioretti, D. L., Simon, G., Saunders, K., ... & Kessler, R. (2005). Chronic spinal pain and physical–mental comorbidity in the United States: results from the national comorbidity survey replication. *Pain, 113*(3), 331-339.

19. Sadosky, A. B., DiBonaventura, M., Cappelleri, J. C., Ebata, N., & Fujii, K. (2015). The association between lower back pain and health status, work productivity, and health care resource use in Japan. *Journal of pain research, 8*, 119.

20. Cedraschi, C., Luthy, C., Allaz, A. F., Herrmann, F. R., & Ludwig, C. (2016). Low back pain and health-related quality of life in community-dwelling older adults. *Eur Spine J, 25*(9), 2822-2832, doi:10.1007/s00586-016-4483-7.

21. Chiarotto, A., Maxwell, L. J., Ostelo, R. W., Boers, M., Tugwell, P., & Terwee, C. B. (2019). Measurement Properties of Visual Analogue Scale, Numeric Rating Scale, and Pain Severity Subscale of the Brief Pain Inventory in Patients With Low Back Pain: A Systematic Review. *J Pain, 20*(3), 245-263, doi:10.1016/j.jpain.2018.07.009.

22. Twisk, J. W. R. (2013). *Applied Longitudinal Data Analysis for Epidemiology*. Cambridge: Cambridge University Press.

23. Maas, E. T., Juch, J. N., Groeneweg, J. G., Ostelo, R. W., Koes, B. W., Verhagen, A. P., et al. (2012). Cost-effectiveness of minimal interventional procedures for chronic mechanical low back pain: design of four randomised controlled trials with an economic evaluation. *BMC Musculoskelet Disord, 13*(1), 260, doi:10.1186/1471-2474-13-260.

24. Juch, J. N., Maas, E. T., Ostelo, R. W., Groeneweg, J. G., Kallewaard, J.-W., Koes, B. W., et al. (2017). Effect of radiofrequency denervation on pain intensity among patients with chronic low back pain: the Mint randomized clinical trials. *Jama*, *318*(1), 68-81.

25. Soer, R., Reneman, M. F., Speijer, B. L., Coppes, M. H., & Vroomen, P. C. (2012). Clinimetric properties of the EuroQol-5D in patients with chronic low back pain. *Spine J*, *12*(11), 1035-1039, doi:10.1016/j.spinee.2012.10.030.

26. Lamers, L. M., Stalmeier, P. F., McDonnell, J., Krabbe, P. F., & van Busschbach, J. J. (2005). [Measuring the quality of life in economic evaluations: the Dutch EQ-5D tariff]. *Ned Tijdschr Geneeskd*, *149*(28), 1574-1578.

27. Whitehead, S. J., & Ali, S (2010). Health outcomes in economic evaluation: the QALY and utilities. *British medical bulletin*, *96*(1), 5-21.

28. Goossens, M. E., Rutten-van Molken, M. P., Vlaeyen, J. W., & van der Linden, S. M. (2000). The cost diary: a method to measure direct and indirect costs in cost-effectiveness research. *J Clin Epidemiol*, *53*(7), 688-695.

29. Hakkaart-van Roijen, L., Van der Linden, N., Bouwmans, C., Kanters, T., & Tan, S. (2015). Costing manual: Methodology of costing research and reference prices for economic evaluations in healthcare. *Diemen, the Netherlands: Zorginstituut Nederland.*

30. Koopmanschap, M. A. (2005). PRODISQ: a modular questionnaire on productivity and disease for economic evaluation studies. *Expert Rev Pharmacoecon Outcomes Res*, *5*(1), 23-28, doi:10.1586/14737167.5.1.23.

31. Goossens, M. E., Rutten-van Mölken, M. P., Vlaeyen, J. W., & van der Linden, S. M. (2000). The cost diary: a method to measure direct and indirect costs in cost-effectiveness research. *J Clin Epidemiol*, *53*(7), 688-695.

32. Ferreira-Valente, M. A., Pais-Ribeiro, J. L., & Jensen, M. P. (2011). Validity of four pain intensity rating scales. *Pain, 152*(10), 2399-2404, doi:10.1016/j.pain.2011.07.005.

33. Williamson A, H. B. (2005). Pain: A review of three commonly used pain rating scales. *J Clin Nurs., 14*(7), 798-804.

34. Childs, J. D., Piva, S. R., & Fritz, J. M. (2005). Responsiveness of the numeric pain rating scale in patients with low back pain. *Spine (Phila Pa 1976), 30*(11), 1331-1334.

35. Vianin, M. (2008). Psychometric properties and clinical usefulness of the Oswestry Disability Index. *J Chiropr Med, 7*(4), 161-163, doi:10.1016/j.jcm.2008.07.001.

36. Roland, M., & Fairbank, J. (2000). The Roland-Morris Disability Questionnaire and the Oswestry Disability Questionnaire. *Spine (Phila Pa 1976), 25*(24), 3115-3124.

37. Fairbank, J. C., & Pynsent, P. B. (2000). The Oswestry Disability Index. *Spine (Phila Pa 1976), 25*(22), 2940-2952; discussion 2952.

38. Walsh T.L., H. B., Lurie J.D., Weinstein J.N. (2003). Is a condition-specific instrument for patients with low back pain/leg symptoms really necessary? The responsiveness of the Oswestry Disability Index, MODEMS, and the SF-36. . *Spine, 28*(6), 607–615.

39. Ostelo, R. W., & de Vet, H. C. (2005). Clinically important outcomes in low back pain. *Best Pract Res Clin Rheumatol, 19*(4), 593-607, doi:10.1016/j.berh.2005.03.003.

40. Chiarotto, A., Boers, M., Deyo, R. A., Buchbinder, R., Corbin, T. P., Costa, L. O. P., et al. (2018). Core outcome measurement instruments for clinical trials in nonspecific low back pain. *Pain, 159*(3), 481-495, doi:10.1097/j.pain.0000000000001117.

41. van Hooff, M. L., van Loon, J., van Limbeek, J., & de Kleuver, M (2014). The Nijmegen decision tool for chronic low back pain.

Development of a clinical decision tool for secondary or tertiary spine care specialists. *PloS one*, *9*(8).

42. Devilly, G. J., & Borkovec, T. D. (2000). Psychometric properties of the credibility/expectancy questionnaire. *J Behav Ther Exp Psychiatry*, *31*(2), 73-86.

43. Davidson, M., & Keating, J. (2005). Oswestry Disability Questionnaire (ODQ). *Aust J Physiother*, *51*(4), 270.

44. Rabin, R., de Charro, F. EQ-5D: a measure of health status from the EuroQol Group. *Ann Med*. 2001 Jul;33(5):337-43.

45. Brazier, J. E., Harper, R., Jones, N. M., O'Cathain, A., Thomas, K. J., Usherwood, T., et al. (1992). Validating the SF-36 health survey questionnaire: new outcome measure for primary care. *Bmj*, *305*(6846), 160-164, doi:10.1136/bmj.305.6846.160.

46. Hays, R. D., & Morales, L. S. (2001). The RAND-36 measure of health-related quality of life. *Ann Med*, *33*(5), 350-357.

47. Van der Zee KI, S. R., Heyink JW, de Haes H (1996). Psychometric qualities of the RAND 36-Item Health Survey 1.0: a multidimensional measure of general health status. *Int J Behav Med*, *3*(2), 104.

48. Lousberg, R., Van Breukelen, G. J., Groenman, N. H., Schmidt, A. J., Arntz, A., & Winter, F. A. (1999). Psychometric properties of the Multidimensional Pain Inventory, Dutch language version (MPI-DLV). *Behav Res Ther*, *37*(2), 167-182.

49. McKillop, J. M., & Nielson, W. R. (2011). Improving the usefulness of the Multidimensional Pain Inventory. *Pain Res Manag*, *16*(4), 239-244.

50. Burton A, B. L., Bryan S. (2007). Cost-effectiveness in clinical trials: using multiple imputation to deal with incomplete cost data. *Clinical Trials*, *4*(2), 154-161.

51. White, I. R., Royston, P., & Wood, A. M. (2011). Multiple

imputation using chained equations: issues and guidance for practice. *Statistics in medicine, 30*(4), 377-399.

52. Luo N, J. J., Coons SJ. (2010). Using instrument-defined health state transitions to estimate minimally important differences for four preference-based health-related quality of life instruments. *Med Care, 48*(4), 365-371.

53. Walters, S. J., & Brazier, J. E. (2005). Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. *Qual Life Res, 14*(6), 1523-1532.

54. Stefane, T., Santos, Amanda Munari dos, Marinovic, Adriano, & Hortense, Priscilla (2013). Chronic low back pain: pain intensity, disability and quality of life. *Acta Paulista de Enfermagem, 26*(1), 14-20.

55. Lamé, I. E., Peters, M. L., Vlaeyen, J. W., Kleef, M. V., & Patijn, J. (2005). (2005). Quality of life in chronic pain is more associated with beliefs about pain, than with pain intensity. *European Journal of Pain, 9*(1), 15-24.

56. Sullivan, M. J., Bishop, S. R., & Pivik, J. (1995). The pain catastrophizing scale: development and validation. *Psychological assessment, 7*(4), 524-532.

57. Picavet, H. S., Vlaeyen, J. W., & Schouten, J. S. (2002). Pain catastrophizing and kinesiophobia: predictors of chronic low back pain. *Am J Epidemiol, 156*(11), 1028-1034, doi:10.1093/aje/kwf136.

58. Henschke, N., Maher, C. G., Refshauge, K. M., Herbert, R. D., Cumming, R. G., Bleasel, J., et al. (2008). Prognosis in patients with recent onset low back pain in Australian primary care: inception cohort study. *Bmj, 337*(a171), a171, doi:10.1136/bmj.a171.

59. Mok, L. C., & Lee, I. F. (2008). Anxiety, depression and pain intensity in patients with low back pain who are admitted to acute care hospitals. *J Clin Nurs, 17*(11), 1471-1480, doi:10.1111/j.1365-2702.2007.02037.x.

60. Linda Muller, J. F., Katja Chkalova, Ben Dankmeyer, Maarten Bloem (2018). Monitor loonverschillen mannen en vrouwen, 2016. https://www.cbs.nl/nl-nl/maatwerk/2018/47/monitor-loonverschillen-mannen-en-vrouwen-2016.

61. Wenig CM, S., C. O., Kohlmann, T., & Schweikert, B (2009). Costs of back pain in Germany. *Eur J Pain 13*(3), 280-286.

62. Herdman, M., Gudex, C., Lloyd, A., Janssen, M. F., Kind, P., Parkin, D., ... & Badia, X (2011). Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Quality of Life Research*, *20*(10), 1727-1736.

63. Janssen, M. F., Pickard, A. S., Golicki, D., Gudex, C., Niewada, M., Scalone, L., ... & Busschbach, J (2013). Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: a multi-country study. *Quality of Life Research*, *22*(7), 1717-1727.

3

Predictive factors of high societal costs among chronic low back pain patients

Elizabeth N. Mutubuki,[1] Mariette A. Luitjens,[1,2] Esther T. Maas,[1,3] Frank J.P.M. Huygen,[4] Raymond W.J.G. Ostelo[1,5] Maurits W. van Tulder,[1,6] Johanna M. van Dongen[1]

[1] Department of Health Sciences, Faculty of Science, Vrije University Amsterdam, Amsterdam Movement Sciences Research Institute, the Netherlands
[2] Child and Adolescent Health Service, Perth, Australia
[3] School of Population and Public Health, University of British Columbia, Canada
[4] Department of Anesthesiology, Centre of Pain Medicine, ErasmusMC, Rotterdam, the Netherlands
[5] Department of Epidemiology and Biostatistics, VU University Medical Center, Amsterdam Movement Sciences Research Institute, the Netherlands
[6] Department of Physiotherapy & Occupational Therapy, Aarhus University Hospital, Aarhus, Denmark

Mutubuki EN, Luitjens MA, Maas ET, et al. Predictive factors of high societal costs among chronic low back pain patients. Eur J Pain. 2020;24(2):325-337. doi:10.1002/ejp.1488

ABSTRACT

Background: Societal costs of low back pain (LBP) are high, yet few studies have been performed to identify the predictive factors of high societal costs among chronic LBP patients. This study aimed to determine which factors predict high societal costs in patients with chronic LBP.

Methods: Data of 6,316 chronic LBP patients were used. In the main analysis, high societal costs were defined as patients in the top 10% of cost outcomes. Sensitivity analyses were conducted using patients in the top 5% and top 20% of societal costs. Potential predictive factors included patient expectations, demographic factors (e.g. age, gender, nationality), socio-economic factors (e.g. employment, education level) and health-related factors (e.g. body mass index [BMI], general health, mental health). The final prediction models were obtained using backward selection. The model's prognostic accuracy (Hosmer–Lemeshow X2, Nagelkerke's R2) and discriminative ability (area under the receiver operating curve [AUC]) were assessed, and the models were internally validated using bootstrapping.

Results: Poor physical health, high functional disability, low health-related quality of life, high impact of pain experience, non-Dutch nationality and decreasing pain were found to be predictive of high societal costs in all models, and were therefore considered robust. After internal validation, the models' fit was good, their explained variance was relatively low (≤14.1%) and their AUCs could be interpreted as moderate (≥0.71).

Conclusion: Future studies should focus on understanding the mechanisms associated with the identified predictors for high societal costs in order to design effective cost reduction initiatives.

Significance: Identifying low back pain patients who are at risk (risk stratification) of becoming high-cost users and making appropriate initiatives could help in reducing high costs.

INTRODUCTION

In recent years, low back pain (LBP) has become the leading cause of years lived with disability in high-, middle- and low income countries (Vos et al., 2017). A 54% increase in years lived with disability caused by LBP was reported worldwide between 1990 and 2015 (Hartvigsen, Hancock, & Kongsted, 2018). Next to the high disease burden of LBP, its economic burden is substantial (Tulder, Koes, & Bombardier, 2002). In 2007, for example, the societal cost of LBP in the Netherlands was estimated to be 3.5 billion euros, which accounted for approximately 0.6% of the Gross National Product (Lambeek et al., 2011). The estimated annual total cost of LBP in the United States is 100 billion dollars, (Dieleman et al., 2016) in Australia 9 billion Australian dollars, (Walker, Muller, & Grant, 2003) in Switzerland 6.6 billion euros (Wieser et al., 2011) and in the UK 12.3 billion British pounds (Maniadakis & Gray, 2000).

A systematic review by Hestbaek, Leboeuf-Yde, and Manniche (2003) showed that in many cases LBP did not resolve on its own and that 62% of LBP patients keep experiencing pain after 12 months (Hestbaek et al., 2003; Verkerk et al., 2013). Nonetheless, the majority of LBP patients do not seek treatment (Ferreira et al., 2010) and Engel, Von Korff, and Katon (1996) and Vlaeyen et al. (2018) reported that it is very likely that the majority of the total societal costs from LBP stem from a relatively small group of chronic LBP patients (Engel et al., 1996; Vlaeyen et al., 2018).

A proactive approach requires identifying high-risk patients accurately before substantial avoidable costs have been incurred and health status has deteriorated. Exploring the mechanisms related to high-cost users could potentially lead to ideas for initiatives or policy measures aimed at reducing costs. A report from The Commonwealth Fund (2012) maintains this view by placing emphasis on the need to address high-cost health care users with chronic conditions if potentially significant gains are to be made (System TCFCoaHPH, 2012). Identifying factors predictive of high societal costs may provide opportunities to create appropriate initiatives aiming to prevent high-cost outcomes as well as result in improvement of patient quality of life and a reduction in health care spending (Buchbinder et al.,

2013; Chechulin, Nazerian, Rais, & Malikov, 2014).

To date, many studies have focused on investigating factors that predict whether acute LBP will become chronic. Various studies have identified a number of predictive factors for LBP chronicity, including high levels of psychological distress, low levels of physical activity, smoking, poor self-rated health and dissatisfaction with employment (Klenerman, Slade, & Stanley, 1995; Linton & Halldén, 1998; Valat, Goupille, & Védere, 1997). However, in practice there is limited success in using this information to prevent or manage chronic LBP (Tulder et al., 2002; Waddell, 2006). Furthermore, whilst predictive factors have commonly been investigated in various other areas of LBP, such as identifying predictive factors for return to work, disability and future health care utilization, few studies have explored the possible factors that are predictive of high societal cost (Becker et al., 2010; Lancourt & Kettelhut, 1992; Pincus, Burton, Vogel, & Field, 2002; Skargren & Öberg, 1998). Therefore, the aim of this study was to identify predictive factors for high societal costs among chronic LBP patients in the Netherlands.

METHODS

Study population and design

A model was constructed to determine factors predicting high societal costs among chronic LBP patients. Data collected during the MinT (minimal invasive treatment) study in the Netherlands were used to develop the model. The MinT study consisted of three randomized controlled trials and an observational study. The aim of the MinT study was to assess the cost-effectiveness of adding minimal interventional procedures to a standardized exercise program, compared with a standardized exercise program alone (Juch et al., 2017; Maas et al., 2012). Patients were eligible for the MinT study in general if they had chronic (>3 months) LBP, showed no improvement of symptoms after conservative treatment, were referred to a pain clinic and were able to complete Dutch questionnaires. Patients were included in the randomized controlled trials and observational study

between 1 January 2013 and 1 July 2014 and between 1 January 2013 and 17 December 2015, respectively. In the present study, only data of the observational study were used. The observational study monitored patients who did not want to, or were not eligible to participate in the aforementioned randomized controlled trials or who received the intervention after recruitment for the randomized controlled trials was closed (between 1 July 2014 and 17 December 2015) (Maas et al., 2012). The exclusion criteria for participating in the randomized controlled trials included, amongst others, patients with a negative diagnostic test, patients with a body mass index (BMI) higher than 35, patients older than 70 years, patients with severe psychiatric or psychological problems, patients diagnosed with facet, disc, sacroiliac (SI) joint or combination pain but did not want to participate in the randomized controlled trials (Maas et al., 2012). The observational data will inform about the proportion of patients with a positive or negative diagnostic test for facet pain, disc pain, SI joint pain and a combination of these, and the clinical outcomes of patients with a negative diagnostic test. Patients diagnosed with facet, disc, SI joint or combination pain, by means of a diagnostic block, will be asked to take part of one of the four RCTs. The observational study will monitor patients who do not want to, or are not eligible to participate in the RCTs.

Ethical approval for the MinT study was obtained from the Medical Ethics Committee of the Erasmus Medical Centre in Rotterdam (registration number MEC-2012–079). Local research governance was obtained from all participating pain clinics and all participants gave written informed consent (Maas et al., 2012).

Outcome measure

The outcome of the current study was having high societal costs (yes/no). Having high societal costs was defined as patients with costs in the top 10th percentile. Previous studies have defined high costs as patients in the top 20–25th percentile (Becker et al., 2010; Engel et al., 1996). A study in the United States studied health care expenditures from 1928 to 1996 found that the top 5% of high-cost users accounted for more than half of health spending, while the top 10% accounted for about 70% of all

health care spending (Berk & Monheit, 2001). For this study, the 10th percentile for societal costs was therefore assumed to be appropriate due to the large sample size.

Societal costs were measured using 3-monthly retrospective cost questionnaires throughout the 1-year study period (i.e. administered at 3-, 6-, 9- and 12-month follow-up; Goossens, Rutten-van Mölken, Vlaeyen, & Linden, 2000). The self-administered cost questionnaires included measures of health care utilization, informal care, unpaid productivity and absenteeism due to back pain. Health care utilization included primary care (e.g. general practitioner care, manual therapy, physical therapy, exercise therapy) and secondary care (e.g. diagnostic and therapeutic interventions, hospitalization). Data from the updated Dutch Manual of Costing
were used to value costs of common health care services (Hakkaart-van Roijen, Van der Linden, Bouwmans, Kanters, & Tan, 2015). For less common health care services, hospital accounting records and/or prices of professional organizations were used. Informal care and unpaid productivity were valued using the recommended Dutch shadow price of €14,32 per hour (Hakkaart-van Roijen et al., 2015). Absenteeism from paid employment was measured using the Productivity and Disease Questionnaire (PRODISQ; Koopmanschap, 2005), and was valued in accordance with the friction cost approach using hourly productivity costs of males and females (Koopmanschap & Rutten, 1996). The friction cost approach assumes that production losses are confined to the period needed to replace a sick worker, which is currently assumed to be 12 weeks in the Netherlands (Hakkaart-van Roijen et al., 2015). All costs were expressed in Euros 2017. An overview of the main cost categories, examples of common sub-cost categories as well as their unit prices can be found in File S1.

Potential predictive factors

Potential predictive factors were based on previous literature (Becker et al., 2010; Chechulin et al., 2014; Engel et al., 1996; Klenerman et al., 1995; Lancourt & Kettelhut, 1992; Linton & Halldén, 1998; Pincus et al., 2002; Skargren & Öberg, 1998; Valat et al., 1997), and measured at baseline and included:

- Treatment credibility and patient expectancy for improvement after treatment (Credibility/Expectancy Questionnaire [CEQ] (Devilly & Borkovec, 2000); scores were transformed to 0—least credibility/expectancy to 100—more credibility/expectancy) to improve comparability of the odds ratios.
- Pain intensity (Numeric Pain Rating Scale [NPRS]; range 0—no pain to 100—worst pain imaginable; Childs, Piva, & Fritz, 2005). Scores were transformed to 0–100 to improve comparability of the odds ratios.
- Functional disability (Oswestry Disability Index [ODI]; range 0—no disability to 100—maximum disability; Davidson & Keating, 2005; Fairbank & Pynsent, 2000).
- Health-related quality of life (EuroQol [EQ-5D-3L]; range 0—worst imaginable health state to 100—best imaginable health state, higher scores indicating better health; Rabin, 2001). The participants' EQ-5D-3L scores were converted into utility scores using the Dutch tariff (Lamers, Stalmeier, McDonnell, & Krabbe, 2005) and the scores were transformed to 0–100 to improve comparability of the odds ratios.
- General health—mental component score and physical component score (Rand-36 [Rand-36]; scores range 0—lowest general health to 100—highest general health) were transformed so that a higher score indicated better health status (Brazier et al., 1992; Hays & Morales, 2001; Vander Zee & Sanderman, 1996). The two dimensions of the Rand-36 form, namely mental and physical health, were entered separately in the model.
- Impact of pain experience (Multidimensional Pain Inventory [MPI]; range 0—least/best to 100 most/worst). Scores were transformed to 0–100 to improve comparability of the odds ratios. For the purpose of this analysis, scores from the five sub-scales of the first section of the MPI were used, that is, pain severity, interference with daily activities, life control, affective distress and support (Lousberg et al., 1999; McKillop & Nielson, 2011).
- Education level low/moderate/high. Low indicates no education, primary level education, lower vocational and lower secondary education; moderate indicates

higher secondary education or undergraduate; high indicates tertiary education, university or postgraduate.
- Body mass index ([BMI], weight in kg/(height in metres)2).
- Employment (yes/no).
- Recurrent complaints (yes/no).
- Age in (years).
- Gender (male/female).
- Nationality (Dutch/non-Dutch).
- Smoking (yes/no).
- Type of health care insurance (basic/additional).
- Region of residence (south/north/east/west).
- Married/living together (yes/no).
- Diagnosis (sacroiliac joint (SI)/facet/disc/combined/unclear). Diagnosis was based on medical history and clinical examination. Both followed a standard format and were performed by experienced clinicians. Depending on the suspected source of pain, clinical examination included provocation tests (compression test; distraction test; FIexion, Abduction, and External Rotation [FABER] test; Gaenslen test; thigh thrust test; Gillett test) and diagnostic anaesthetic blocks. For a more detailed description of the diagnostic procedures, we refer elsewhere (Juch et al., 2017; Maas et al., 2012).

Statistical analysis

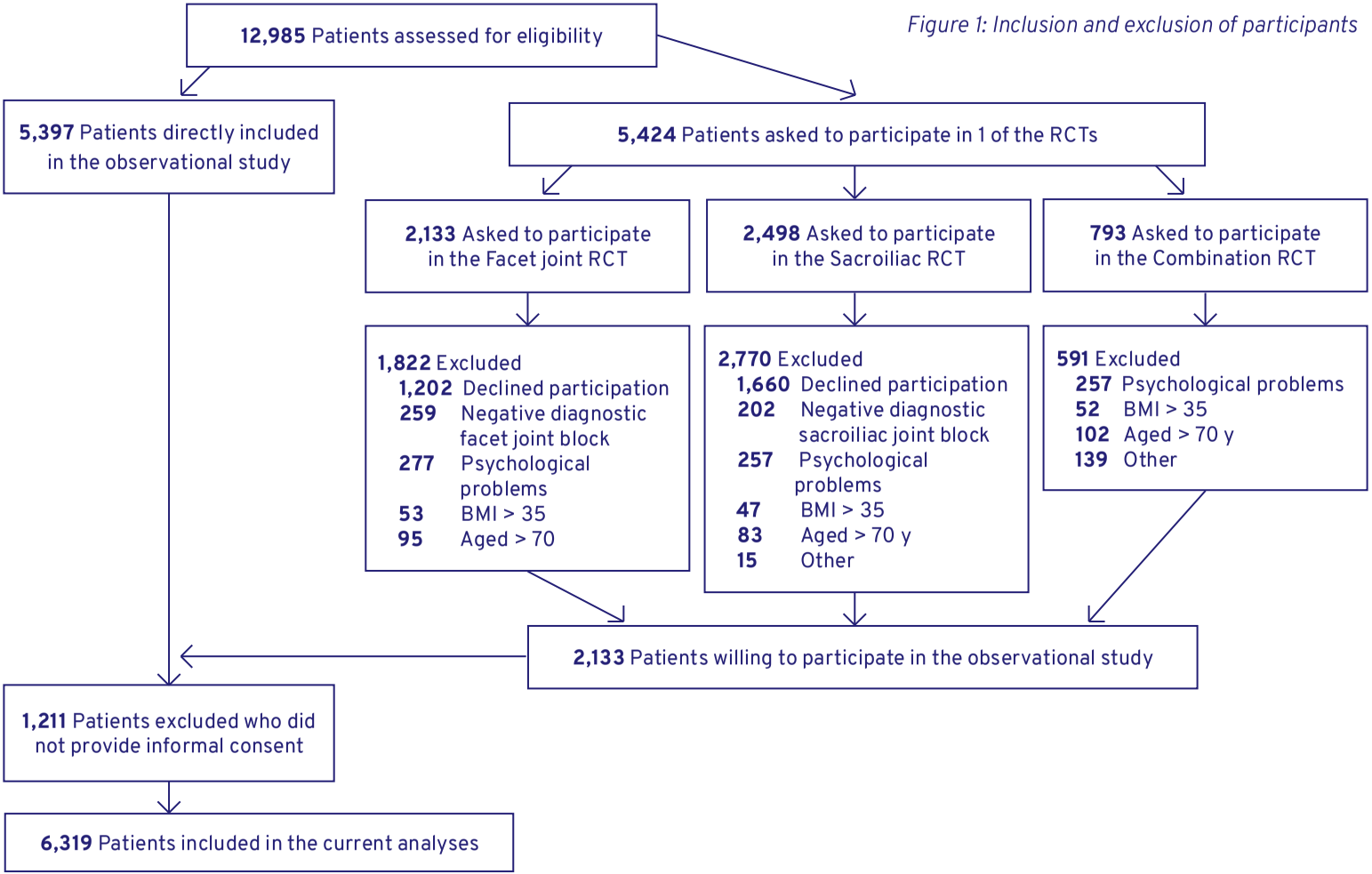
The prediction model was constructed using multivariable logistic regression analysis (Harrell, Lee, & Mark, 1996; Steyerberg, 2010). Prior to constructing the model, missing data were handled using multiple imputation to avoid possible bias due to selective drop-out of participants, which might influence the results when conducting a complete-case analysis (Burton, Billingham, & Bryan, 2007). Imputations were performed by treatment group and per time point using predictive mean matching. Following this, tests were conducted to verify the linearity and additivity assumptions (Harrell et al., 1996).

Manual backward selection was used to obtain the final predictive factors with a $p < .10$. Variables with the highest p-value were excluded from the model one by one and the

analysis was rerun until only variables with a $p < .10$ constituted the model. A $p < .10$ was used to ensure that predictions are accurate, whilst preventing type-1 errors caused by overfitting (Harrell et al., 1996). The overall performance and predictability of the model were tested using Nagelkerke's R2 (Bewick, Cheek, & Ball, 2005; Greiner, Pfeiffer, & Smith, 2000; Steyerberg et al., 2010). Other performance measures included the area under the "receiver operating characteristics" (ROC) curve to measure the final model's discriminative value (area under the receiver operating curve [AUC]) (Bewick et al., 2005; Greiner et al., 2000; Steyerberg et al., 2010) as well as the Hosmer–Lemeshow goodness-of-fit to measure the calibration of the model (Bewick et al., 2005; Greiner et al., 2000; Steyerberg et al., 2010). To adjust for the fact that the model was developed and tested in the same population, which typically causes regression coefficients and performance measures to be overestimated (i.e. overfitting), bootstrapping was used to internally. validate the model (Bewick et al., 2005; Greiner et al., 2000; Steyerberg et al., 2010). Multiple imputation and multivariate regression analyses were conducted using Stata (version 14SE, Stata Corp), and internal validation was performed using R (i386 version 3.1.2).

To test the robustness of the results, two sensitivity analyses were conducted; (a) using the top 20th percentile for high costs, and (b) using the top 5th percentile for high costs.

Figure 1: Inclusion and exclusion of participants



RESULTS

Participants

Data from 6,316 chronic LBP patients in the observational study group were analysed in the present study (Figure 1). Of them, the majority were female (66%), overweight (67%), Dutch (95%), had a low level of education (56%) and more than half were unemployed (59%; Table 1). Most of the predictive factors had about 17% of patients with missing data. The amount of missing values for all the variables entered in the model are reported in File S2. Costs at different cut-off points were as follows: 10% (≥€11,922), 5% (≥€19,403) and 20% (≥€7,906). The average societal costs per patient were €5,522 and the median costs were €2,995.

Development, performance and internal validity of the top 10% prediction model

Females, non-Dutch nationals, combined diagnosis (LBP caused by both facet joints and intervertebral disc), poor physical health, high functional disability, low health-related quality of life, decreasing age, high impact of pain experience and decreasing pain intensity were found to increase the odds of having high societal costs (Table 2). The Hosmer–Lemeshow statistic was not significant ($X2 = 7$, $p = .55$), indicating that the model's overall fit was good. The model explained 14.3% (Nagelkerke's R2) of the variation in the outcome (i.e. high societal costs) and the model's AUC was 0.74 (95% CI 0.67–0.72). After internal validation, the model's explained variance was 13.2% and the AUC was 0.73. The calibration slope was 0.97, indicating relatively little optimism or overfitting of the regression coefficients.

Sensitivity analysis

Using an outcome consisting of patients in the top 20th percentile of societal costs, combined diagnosis, poor physical health, high functional disability, low health-related quality of life, high impact of pain experience, non-Dutch nationality, decreasing pain intensity and being female were predictive factors of having

high societal costs (Table 3). The Hosmer–Lemeshow statistic was not significant (X2= 8.5, p= .47), Nagelkerke's R2 was 0.146 and the model's AUC was 0.72. After internal validation, the model's explained variance reduced to 14.1% and the AUC to 0.71. The calibration slope was 0.98.

Using an outcome consisting of patients in the top 5th percentile of societal costs, high-level education, poor physical health, high functional disability, low health-related quality of life, high impact of pain experience, non-Dutch nationality and decreasing pain intensity were predictive factors of having high societal costs (Table 4). The Hosmer–Lemeshow statistic was not significant (X2 = 7.2, p = .59), indicating that the model's overall fit was good. The model explained 14.1% (Nagelkerke's R2) of the variation in the outcome (high costs) and the model's AUC was 0.76. After internal validation, the model's explained variance reduced to 13.2% and the AUC to 0.76. The calibration slope was 0.97.

Table 5 provides an overview of robust predictors of high societal costs in all three models

DISCUSSION

Main findings

High impact of pain experience (MPI interference), being female, non-Dutch national, combined diagnosis (LBP caused
by both facet joints and intervertebral disc), poor physical health, high functional disability, low health-related quality of life, younger age and decreasing pain intensity were found to increase the odds of having high societal costs. The model's overall fit was good and its explained variance was relatively low (Bewick et al., 2005; Greiner et al., 2000; Steyerberg et al., 2010) that is, only 14.3% of the variance in high societal costs was explained by the identified predictive factors. The AUC was 0.73 and can be interpreted as moderate (Greiner et al., 2000). Internal validation had little effect on the model's performance, illustrating minimal chance of overfitting of the regression coefficients (Steyerberg et al., 2010).

At a 5% cut-off point in our sensitivity analysis, high education level became a predictor and gender and age were no longer predictors. There were no additional predictive factors when a cut-off point of 20% was used, instead age was no longer a predictor. The performance of the sensitivity analyses models was equal to that of the main analysis. Poor physical health, high functional disability, low health-related quality of life, high impact of pain experience, non-Dutch nationality and decreasing pain were found to be predictive of having high societal costs in all models, suggesting that they constitute the most robust predictors of high societal costs.

Comparison with literature

Few studies have focused on investigating predictive factors for high societal costs among chronic LBP patients. A study by Engel et al. (1996) reported increasing chronic pain grade and pain persistence as strong predictors of high costs and high back pain costs, followed by disc disorder/ sciatica diagnosis and increasing depressive symptoms. Diagnosis as a predictor of high costs is in line with the results of the present study as well as those of previous ones (Becker et al., 2010; Wenig, Schmidt, Kohlmann, & Schweikert, 2009). In contrast to the present study, they found mental health and high pain scores to be predictors for high costs. Mental health was also a predictor of high societal costs in the studies of Becker et al. (2010) and Ritzwoller, Crounse, Shetterly, and Rublee (2006). This discrepancy could be due to different cut-off points for high costs (>20% in the previous studies vs. 10% in the present study). The definition of mental health (i.e. depression vs. general mental health) varied among the studies, Becker et al. (2010) focused on depression, whereas Ritzwoller et al. (2006) included anxiety, depression and psychosis. Differences in measuring mental health were noted, 1-item question (present study) versus a risk adjustment system used to identify comorbidities (Ritzwoller et al., 2006) versus CES-D ranging from 0 to 60 (Becker et al., 2010). Depression was associated with high health care costs in the study of Becker et al. (2010) and a possible explanation was that physicians initiate costly health care when confronted with mood disorders (Becker et al., 2010). Ritzwoller et al. (2006) reported an association of

depression and psychopathy with increased LBP episodes and high costs. Comorbidities have been associated with longer duration of LBP and work disabilities (Nordin et al., 2002).

Although previous studies have reported an increase in LBP intensity to be a predictor of high costs (Becker et al., 2010; Wenig et al., 2009), the present study reported decreasing pain intensity as a predictor of high costs. A possible explanation for this discrepancy is that only chronic LBP was included in the present study versus general LBP (acute and chronic; Becker et al., 2010; Ekman, Jönhagen, Hunsche, & Jönsson, 2005; Engel et al., 1996; Ritzwoller et al., 2006; Wenig et al., 2009) and that the studies took place in different health care settings, that is, primary (Becker et al., 2010; Ekman et al., 2005; Engel et al., 1996; Ritzwoller et al., 2006) versus secondary (present study). Fink-Miller, Long, and Gross (2014) reported that chronic LBP patients in primary care reported more severe pain compared to chronic LBP patients in tertiary care and suggest shorter duration of complaints and shopping for opioids by chronic LBP patients in primary care as possible explanations (Fink-Miller et al., 2014). Also, patients presenting in secondary and/or tertiary care may have exhausted conservative therapies, hence could have already made high costs.

Contrary to the findings of Wenig et al. (2009), being female was a predictor of high costs in the present study and in previous studies (Ekman et al., 2005). Wenig et al. (2009) reported that women had a higher probability to cause high costs and utilized health care more quickly than men and when men used health care for LBP it resulted in higher costs on average. The present study had almost double the amount of women compared to men, whereas there was a small difference in the amount of men and women in the study of Wenig et al. (2009).

Another important difference between the present study and the previous ones is the applied perspective. In the present study, a societal perspective was applied, including health care, absenteeism, informal care and unpaid productivity costs, whereas Engel et al. (1996) and Ritzwoller et al. (2006) only included health care costs. Becker et al. (2010) evaluated costs from a societal perspective but did not include informal care costs, Wenig et al. (2009) also applied a societal approach that included health care and lost productivity costs, but did not

include informal care costs.

Also important to note is the higher Nagelkerke's R2 for the model by Becker et al. (i.e. 0.28) compared to that of the present study (i.e. 0.14). Information regarding the fit of the model (Nagelkerke's R2, AUC) is missing from some previous studies (Engel et al., 1996; Wenig et al., 2009). In the present study, the explained variance was probably lower than that of other studies because we applied the broadest perspective, that is, the societal one. The relatively low explained variance may also be interpreted as the variables entered into our model are less suitable at predicting high costs (Ekman et al., 2005), important predictors are missing or chronic LBP patients who are having high costs are a heterogeneous population. Demographic, social and clinical factors included in this model, as in other prediction studies, are typically measured in LBP studies.

Other predictive factors of high costs include diabetes, rheumatoid arthritis, back pain persistence (Engel et al., 1996), fear of avoidance beliefs (Becker et al., 2010), low education and unemployment (Wenig et al., 2009). In contrast to our findings, both low education level and unemployment were not predictors of high costs in our sensitivity analysis, but high education level was. A possible explanation for this is that, 86% of the patients included in this study had comprehensive health care insurance. Highly educated persons are likely to afford more expensive and comprehensive insurance packages offering more options for health care and visits to alternative medicine and therapies. This finding has important
implications for the understanding of the relation between socio-economic status and high-cost users in chronic LBP. In addition, for interventions and policies aimed at highly educated high-cost users in LBP.

In the present study, the average societal costs per patient were €5,522, whereas Dutmer et al (2019) reported around €9,000 in societal costs per patient (Dutmer et al., 2019). This difference could have resulted from the absence of presenteeism costs in the present study, whereas Dutmer et al (2019) did include this cost category in their societal cost estimation. As a consequence, some productivity costs may have been missed. In addition, only patients from a secondary setting were included in the present study, whereas Dutmer et al (2019) included

patients from both secondary and tertiary settings. Tertiary settings are generally more costly compared to secondary settings. Moreover, Dutmer et al (2019) reported higher levels of disability than were reported in the present study, while high levels of disability are typically associated with high costs in LBP (Hartvigsen et al., 2001; Lambeek et al., 2011).

Strength and limitations

Strengths of the present study include that it was one of the very few studies to identify predictive factors for high costs in patients with chronic LBP and that the societal perspective was applied. The large cohort of observed patients with chronic LBP (n = 6,316) greatly increases the power of this study and improves sensitivity to weak predictive factors. Imputation methods were used to deal with missing data thereby avoiding complete-case analysis which would have significantly reduced the power of these findings and potentially introduced information bias due to selective drop-out of participants. Multiple imputation is the preferred statistical method for dealing with missing data, particularly when costs are involved (Burton et al., 2007). Furthermore, internally validating the model by bootstrapping with 250 replications improved the generalizability and robustness of these findings (Bewick et al., 2005; Steyerberg et al., 2013).

Some limitations are notable as well. Although mainly valid and reliable questionnaires were used, the predictive factors were measured using self-reported questionnaires and this might have caused recall and or social desirability bias. Second, presenteeism costs were not included in our analyses, whereas presenteeism has previously been found to be a very important cost driver and is increasingly being recognized as an important problem in the occupational setting (Tsuboi, Murata, Naruse, & Ono, 2019). Hence, further productivity losses could have been missed. Future studies should therefore include presenteeism costs. Third, there is no consensus regarding the most ideal cut-off point for defining high costs. Although in this study different cut-off points, that is, 10% (≥€11,922), 5% (≥€19,403) and 20% (≥€7,906), were used to assess the robustness of the model, a consensus should be reached on the definition of high costs. This

will enable the results to be more comparable and also determine the most suitable moment for initiatives aiming to reduce these costs to be applied. Fourth, in spite of the relatively large sample size of the current study (n = 6,316), there were some predictive factors for which there were very few participants. For example, there were only four (2.8%) non-Dutch nationals in the high-cost group in the main analysis, and it is unknown whether these four participants are representative of all non-Dutch LBP patients. As a consequence, even though non-Dutch nationality was identified as a predictor in all of the models, further research is needed to establish whether non-Dutch nationality is indeed a very strong predictor of having high societal costs among LBP patients. Fifth, the secondary care setting of this study may to some extent limit the generalizability of its findings to other types of LBP patients and/or other settings. Amongst others, the relatively high unemployment rate of 59% may have resulted in an underestimation of the productivity costs, whereas secondary care is generally more expensive than primary care and health care costs may thus have been overestimated (Lambeek et al., 2011). As a consequence, the total societal cost estimates are likely to be specific to the secondary care setting. Furthermore, the disability rate in this study is rather low in comparison to other studies conducted
in secondary settings (Dutmer et al., 2019), therefore caution should be exercised when applying these results to other populations. Sixth, apart from high BMI-related diseases no other comorbidities have been included in the study. Overweight and obesity are well represented in the present study because these were exclusion criteria for the RCTs in the Mint study.

Implications for research and practice

The lack of professional consensus regarding a cut-off point for high costs is probably due to limited studies in this field. Having a consensus regarding a cut-off point can enable comparisons to be made and it is essential in policy and decision making. Identifying those patients who are at risk (risk stratification) of becoming high-cost users and making appropriate initiatives could help in reducing high costs. For example, non-Dutch nationality might be associated with a more limited mastery of

the language. Maybe the information provided to non-Dutch patients should be adapted. Functional disability and poor physical health are predictors of high societal costs, therapies targeting limitations in activities could play a role in reducing societal costs. There is evidence from randomized controlled trials that stratified care models limit long-term disability arising from LBP (Linton, Nicholas, & Shaw, 2018). These considerations have important implications for how the link between socio-economic status and high-cost use is understood and for policies and programs targeting high-cost use.

CONCLUSION

The present study identifies patients at risk of becoming high cost users and future studies should focus on understanding the mechanisms associated with the identified predictors for high-cost users in order to be able to design and tailor effective cost reduction initiatives.

AUTHOR CONTRIBUTIONS

EM wrote the initial version of the manuscript. EM, ML, JvD and ET were involved in the data analysis process. All authors discussed the results and commented on the manuscript. FH, MvT and RO received funding for the study.

Table 1: Patients' characteristics, all patients and according to societal costs (high versus low)

Participant characteristic	All patients (n=6316)	High costs (n=171)	Low costs (n=6145)
Age (years) [mean (SD)]	57.2 (13,4)	57.6 (12.0)	57.2 (13.5)
Gender [n (%)]			
Female	4142 (66)	128 (75)	4014 (66)
Male	2093 (34)	43 (25)	2050 (34)
BMI [n (%)]			
BMI<18.5 (underweight)	37 (1)		37 (1)
BMI≥18.5<25 (normal weight)	1687 (32)	42 (26)	1645 (32)
BMI ≤25<30 (overweight)	2060 (39)	62 (38)	1998 (39)
BMI≥30 (obese)	1463 (28)	60 (37)	1403 (28)
Smoking [n (%)]			
Yes	1413 (26)	42 (25)	1371 (26)
No	3920 (73)	125 (75)	3795 (73)
Educational level [n (%)]			
Low *(no education, primary level education, lower vocational and lower secondary education)*	2925 (56)	100 (62)	2825 (56)
Moderate *(higher secondary education or undergraduate)*	1467 (28)	43 (27)	1424 (28)
High *(tertiary, university level, postgraduate)*	830 (16)	19 (12)	811 (16)
Living together with a partner [n (%)]			
Yes	4663 (75)	135 (79)	4528 (74)
No	1593 (26)	36 (21)	1557 (26)
Nationality [n (%)]			
Dutch	5049 (95)	163 (98)	4886 (95)
Non-Dutch:	278(5.2)	4(2.8)	274(5.3)
Surinamese	*21 (0.4)*	*0*	*21 (0.4)*
Antillean/Aruban	*22 (0.4)*	*0*	*22 (0.4)*
Turkish	*63 (1)*	*1 (1)*	*62 (1)*
Moroccan	*42 (1)*	*0*	*42 (1)*
Other	*130 (2.4)*	*3 (1.8)*	*127 (2.5)*
Region in the Netherlands [n (%)]			
South	2029 (32)	59 (35)	1970 (32)
North	1165 (19)	30 (18)	1135 (19)
East	1280 (20)	43 (25)	1237 (20)
West	1782 (28)	39 (23)	1743 (29)
Employment [n (%)]			
Yes	1687(42)	66 (39)	1621 (42)
No	2376 (59)	105 (61)	2271 (58)
Recurrent low back pain [n (%)]			
Yes	3,174 (63)	101 (62)	3073 (63)
No	1876 (37)	61 (38)	1815 (37)

Continued - Table 1: Patients' characteristics, all patients and according to societal costs (high versus low)

Diagnosis-source of pain [n (%)]			
1= SI	1864 (33)	57(36)	1807 (33)
2= Facet	2269 (41)	54 (34)	2215 (41)
3= Disc	18 (0.3)	1 (0.63)	17 (0.3)
4= Combined	1391 (25)	44 (28)	1347 (25)
5= Unclear	66 (1)	3 (2)	63 (1)
Patients expectations			
Credibility [mean (SD)] range 0-100	77.1 (17.5)	77 (19.1)	77.1 (17.3)
Expectancy [mean (SD)] range 0-100	57.8 (17.3)	57.2 (17)	58 (17)
Rand-36			
Mental [mean (SD)] range 0-100	22.6 (5)	21.6 (5)	22.6 (5)
Physical [mean (SD)] range 0-100	18.5 (4)	16.0 (4)	18.6 (4)
Health related quality of life(utility) [mean (SD)] range 0-100	48 (29)	31 (28)	48 (29)
MPI [mean (SD)] range per subscale 0-100			
Pain severity	22.6 (5.7)	25.4 (4.4)	22.5 (5.7)
Interference with daily activities	5.8 (1.9)	6.9 (1.6)	5.8 (1.9)
Life control	21.2 (6.3)	20.2 (7)	21.2 (6.2)
Affective distress	15.4 (4.6)	16.5 (4.9)	15.3 (4.6)
Support	28.6 (7.6)	30.4 (6.2)	28.4 (7.6)
Type of health care insurance [n (%)]			
Basic insurance	633 (12)	14 (8)	619 (12)
Comprehensive (basic+additional cover)	4630 (86)	153 (92)	4477 (86)
I don't know	55 (1)	0	55 (1)
ODI functional disability [mean (SD)] range 0-100	11.1 (9)	17.1 (10)	11.1 (9)
Pain intensity [mean (SD)] range 0-100	73 (16)	77 (14)	73 (16)

Note: percentages have been rounded off hence values a bit less than 100% and a bit more that 100%
Scores for MPI, Rand 36, patient expectations, health related quality of life were transformed to a range of 0-100 to enable comparability with the odds ratio.
Diagnosis was based on patient history and physical examination
ODI-Oswestry Disability Index
MPI-Multidimensional Pain Inventory

Table 2: Multivariate model using the top 10th percentile of societal costs as an outcome

	[a]**Coefficient** (regression)	[b]**SE** (of regression coefficient)	**P-value**	[c]**95%CI**		
				Lower bound	Odds Ratio	Upper bound
Diagnosis (ref: sacroiliac joint)						
Facet	0.097	0.139	0.487	0.836	1.102	1.452
Disc	0.109	0.983	0.912	0.161	1.115	7.744
Combined	0.263	0.142	0.066	0.982	1.301	1.725
Unclear	0.731	0.442	0.100	0.868	2.077	4.972
Physical health (Rand-36); range 0-100	-0.069	0.021	0.002	0.895	0.933	0.973
Functional disability (ODI); range 0-100	0.035	0.008	0.000	1.019	1.036	1.053
Health related quality of life (EQ-5D-3L) ; range 0-100	-0.006	0.029	0.052	0.989	0.994	1.000
Impact of pain experience (MPI interference) range0-100	0.188	0.051	0.000	1.092	1.207	1.336
Nationality (ref: non-Dutch)	-0.818	0.215	0.000	0.286	0.441	0.680
Pain intensity (NPRS) range 0-100	-0.011	0.004	0.010	0.981	0.989	0.997
Age (years)	-0.009	0.004	0.031	0.982	0.991	0.999
Gender (ref: female)	-0.214	0.111	0.055	0.649	0.807	1.004
Constant	-0.392	0.762	0.608	0.148	0.676	3.080

[a] *Coefficient multivariable logistic regression*
[b] *SE-standard error*
[c] *CI-confidence interval*

Table 3: Multivariate model using the top 20th percentile of societal costs as an outcome

	[a]**Coefficient** (regression)	[b]**SE** (regression coefficient)	**P-value**	[c]**95%CI**		
				Lower bound	Odds Ratio	Upper bound
Diagnosis (ref: sacroiliac joint)						
Facet	0.053	0.113	0.640	0.841	1.054	1.322
Disc	0.355	0.722	0.624	0.342	1.426	5.954
Combined	0.280	0.110	0.013	1.063	1.323	1.649
Unclear	0.297	0.378	0.433	0.638	1.346	2.843
Physical health (Rand-36); range 0-100	-0.056	0.014	0.000	0.919	0.946	0.973
Functional disability (ODI); range 0-100	0.028	0.007	0.000	1.015	1.028	1.043
Health related quality of life (EQ-5D-3L) ; range 0-100	-0.005	0.002	0.006	0.991	0.995	0.998
Impact of pain experience (MPI interference) range 0-100	0.176	0.036	0.000	1.110	1.192	1.280
Nationality (ref: non-Dutch)	-0.948	0.226	0.000	0.244	0.388	0.616
Pain intensity (NPRS); range 0-100	-0.010	0.003	0.002	0.984	0.990	0.996
Gender (ref: female)	-0.239	0.089	0.008	0.660	0.787	0.939
Constant	-0.027	0.559	0.962	0.318	0.973	2.983

[a] *Coefficient multivariable logistic regression*
[b] *SE-standard error*
[c] *CI-confidence interval*

Table 4: Multivariate model using the top 5th percentile of societal costs as an outcome

	[a]**Coefficient** (regression)	[b]**SE** (regression coefficient)	**P-value**	[c]**95%CI**		
				Lower bound	Odds Ratio	Upper bound
Education (ref: low)						
Medium	0.099	0.201	0.626	0.738	1.104	1.650
High	0.396	0.227	0.086	0.940	1.486	2.336
Physical health (Rand-36); range 0-100	-0.078	0.026	0.004	0.878	0.925	0.974
Functional disability (ODI) ; range 0-100	0.041	0.011	0.000	1.019	1.042	1.065
Health related quality of life (EQ-5D-3L) ; range 0-100	-0.008	0.003	0.028	0.986	0.992	0.999
Impact of pain experience (MPI interference) range 0-100	0.183	0.063	0.004	1.061	1.201	1.359
Nationality (ref: non-Dutch)	-0.855	0.248	0.001	0.259	0.425	0.697
Pain intensity (NPRS) ; range 0-100	-0.013	0.006	0.023	0.975	0.987	0.998
Constant	-1.413	0.904	0.121	0.040	0.243	1.463

[a] *Coefficient multivariable logistic regression*
[b] *SE-standard error*
[c] *CI-confidence interval*

Table 5: Robust predictors of high societal costs in all 3 models

	top 10th percentile	**top 20th percentile**	**top 5th percentile**
	Odds ratio (95% CI)	Odds ratio (95% CI)	Odds ratio (95% CI)
Physical health (Rand-36); range 0-100	0.933 (0.895-0.973)	0.946 (0.919-0.973)	0.926 (0.878-0.976)
Functional disability (ODI); range 0-100	1.036 (1.019-1.053)	1.028 (1.015-1.043)	1.041 (1.018-1.063)
Health related quality of life (EQ-5D-3L) ; range 0-100	0.994 (0.989-1.000)	0.995 (0.991-0.998)	0.992 (0.985-0.999)
Impact of pain experience (MPI interference) range 0-100	1.017 (1.008-1.027)	1.016 (1.010-1.016)	1.017 (1.005-1.028)
Nationality (ref: non-Dutch)	0.441 (0.286-0.680)	0.388 (0.244-0.616)	0.424 (0.258-0.698)
Pain intensity (NPRS); range 0-100	0.989 (0.981-0.997)	0.990 (0.984-0.996)	0.987 (0.976-0.998)

CI-confidence interval

SUPPLEMENTARY FILES

S1. Main cost categories, examples of common sub-cost categories, and unit prices.

Main cost categories	Costs included	Common sub-cost categories	Unit prices (Euros 2017)
Healthcare costs	Primary healthcare costs related to the participants chronic low back pain	General practitioner	€33.76/visit
		Physiotherapy	€33.76/visit
		Social worker	€59.01/visit
		Psychologist	€94.45/visit
		Manual therapist	€38.79/visit
		Chiropractor	€50.37/visit
		Ergotherapy	€33.76/visit
		Haptotherapy	€77.69/visit
		Podotherapy	€60.93/visit
		Accupuncture	€50.37/visit
		Ceasar therapy	€34.79/visit
		Magnetizer	€28.65/visit
		Massage	€42.65/visit
	Secondary health care costs related to the participants chronic low back pain	Hospital admission	€502.26/day
		Day treatment in hospital	€226.63/day
		Policlinic visits	€81.85/visit
		Intensive care treatment	€2399.20/day
		Community/home care	€26.45/hour
		Radiofrequency denervation	Facet joint: €927/ treatment Sacroiliac joint: €815/treatment
		Diagnostic block	€253/treatment
Informal care costs	Costs related to all hours of care provided by family, friends, and other volunteers due to the participants' chronic low back pain		€14.32/hour

Unpaid productivity costs	Costs related to all hours of volunteer work, domestic and educational activities that participants were not able to perform due to their chronic low back pain	€14.32/hour
Absenteeism costs	Costs related to all hours absence from work due to the participants' chronic low back pain	Male: €38.78/ hour Female: €32.33/ hour

S2. Potential predictive variables and the amount of missing variables.

Variables	Number of missing values	Number of non-missing values	Percentage of missing values
Education	1094	5222	17
Diagnosis	708	5608	11
Patients Expectations	1062	5254	17
SF-36: Mental health	1068	5248	17
SF-36 :Physical health	1063	5253	17
BMI	1069	5247	17
Health related quality of life (utility)	1055	5261	17
MPI: Pain severity	1069	5247	17
MPI: Interference with daily activities	1075	5241	17
MPI: Life control	1075	5241	17
MPI: Affective distress	1075	5241	17
MPI: Support	1071	5245	17
NPRS: Pain	1054	5262	17
Employment	2253	4063	36
Chronic complaints	1070	5246	17
Recurrent complaints	1266	5050	20
Age	982	5334	16
Sex	81	6235	16
Dutch nationality	989	5327	16
Smoking	955	5361	16
Type of health care insurance	955	5361	16
Region in Netherlands	60	6256	1
Marital status	60	6256	1
ODI: Functional disability	1059	5257	20

REFERENCES

Becker, A., Held, H., Redaelli, M., Strauch, K., Chenot, J. F., Leonhardt, C., ...Donner-Banzhoff, N.. (2010). Low back pain in primary care: Costs of care and prediction of future health care utilization. *Spine*, *35*(18), 1714–1720. https ://doi.org/10.1097/ BRS.0b013 e3181
cd656f

Berk, M. L., & Monheit, A. C. (2001). The concentration of health care expenditures, revisited. *Health Affairs*, *20*(2), 9–18. https :// doi.org/10.1377/hltha ff.20.2.9

Bewick, V., Cheek, L., & Ball, J. (2005). Statistics review 14: Logistic regression. *Critical Care*, *9*(1), 112–118.

Brazier, J. E., Harper, R., Jones, N. M., O'Cathain, A., Thomas, K. J., Usherwood, T., & Westlake, L. (1992). Validating the SF-36 health survey questionnaire: New outcome measure for primary care. *BMJ*, *305*(6846), 160–164. https ://doi.org/10.1136/ bmj.305.6846.160

Buchbinder, R., Blyth, F. M., March, L. M., Brooks, P., Woolf, A. D., & Hoy, D. G. (2013). Placing the global burden of low back pain in context. *Best Practice & Research Clinical Rheumatology*, *27*(5), 575–589. https ://doi.org/10.1016/j.berh.2013.10.007

Burton, A., Billingham, L. J., & Bryan, S. (2007). Cost-effectiveness in clinical trials: Using multiple imputation to deal with incomplete cost data. *Clinical Trials*, *4*(2), 154–161. https :// doi.org/10.1177/1740774507 076914

Chechulin, Y., Nazerian, A., Rais, S., & Malikov, K. (2014). Predicting patients with high risk of becoming high-cost healthcare users in Ontario (Canada). *Healthcare Policy*, *9*(3), 68–79. https ://doi. org/10.12927/ hcpol.2014.23710

Childs, J. D., Piva, S. R., & Fritz, J. M. (2005). Responsiveness of the numeric pain rating scale in patients with low back pain.

Spine, *30*(11), 1331–1334. https ://doi.org/10.1097/01.brs.00001 64099.92112.29

Davidson, M., & Keating, J. (2005). Oswestry disability questionnaire (ODQ). *Australian Journal of Physiotherapy*, *51*(4), 270. https ://doi.org/10.1016/S0004-9514(05)70016-7

Devilly, G. J., & Borkovec, T. D. (2000). Psychometric properties of the credibility/expectancy questionnaire. *Journal of Behavior Therapy and Experimental Psychiatry*, *31*(2), 73–86. https ://doi.org/10.1016/S0005-7916(00)00012-4

Dieleman, J. L., Baral, R., Birger, M., Bui, A. L., Bulchis, A., Chapin, A., … Murray, C. J. L. (2016). US spending on personal health care and public health, 1996–2013. *JAMA*, *316*(24), 2627–2646. https ://doi.org/10.1001/jama.2016.16885

Dutmer, A. L., Preuper, H. R. S., Soer, R., Brouwer, S., Bültmann, U., Dijkstra, P. U., … Wolff, A. P. (2019). Personal and societal impact of low back pain: The groningen spine cohort. *Spine*. https ://doi.org/10.1097/BRS.00000 00000 003174

Ekman, M., Jönhagen, S., Hunsche, E., & Jönsson, L. (2005). Burden of illness of chronic low back pain in Sweden: A cross-sectional, retrospective study in primary care setting. *Spine*, *30*(15), 1777–1785. https ://doi.org/10.1097/01.brs.00001 71911.99348.90

Engel, C. C., Von Korff, M., & Katon, W. J. (1996). Back pain in primary care: Predictors of high health-care costs. *Pain*, *65*(2–3), 197–204. https ://doi.org/10.1016/0304-3959(95)00164-6

Fairbank, J. C., & Pynsent, P. B. (2000). The Oswestry disability index. *Spine*, *25*(22), 2940–2953. https ://doi.org/10.1097/00007 632-200011150-00017

Ferreira, M. L., Machado, G., Latimer, J., Maher, C., Ferreira, P. H., & Smeets, R. J. (2010). Factors defining care-seeking in low back pain–A meta-analysis of population based surveys. *European Journal of Pain*, *14*(7), 747.e1–747.e7. https ://doi.org/10.1016/j.

ejpain.2009.11.005

Fink-Miller, E. L., Long, D. M., & Gross, R. T. (2014). Comparing chronic pain treatment seekers in primary care versus tertiary care settings. *The Journal of the American Board of Family Medicine*, *27*(5), 594–601. https ://doi.org/10.3122/jabfm.2014.05.130311

Goossens, M. E., Rutten-van Mölken, M. P., Vlaeyen, J. W., & van der Linden, S. M. (2000). The cost diary: A method to measure direct and indirect costs in cost-effectiveness research. *Journal of Clinical Epidemiology, 53*(7), 688–695. https ://doi.org/10.1016/S0895-4356(99)00177-8

Greiner, M., Pfeiffer, D., & Smith, R. (2000). Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. *Preventive Veterinary Medicine*, *45*(1–2), 23–41. https://doi.org/10.1016/S0167-5877(00)00115-X

Hakkaart-van Roijen, L., Van der Linden, N., Bouwmans, C., Kanters, T., & Tan, S. (2015). Costing manual: Methodology of costing research and reference prices for economic evaluations in healthcare. *Diemen, the Netherlands: Zorginstituut Nederland.*

Harrell, F. E., Lee, K. L., & Mark, D. B. (1996). Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine, 15*(4), 361–387. https ://doi.org/10.1002/(SICI)1097-0258(19960229)15:4<361:AID-SIM16 8>3.0.CO;2-4

Hartvigsen, J., Hancock, M. J., Kongsted, A., et al. (2018). What low back pain is and why we need to pay attention. *The Lancet., 391*(10137), 2356–2367. https ://doi.org/10.1016/S0140-6736(18)30480-X

Hays, R. D., & Morales, L. S. (2001). The RAND-36 measure of healthrelated quality of life. *Annals of Medicine*, *33*(5), 350–357. https ://doi.org/10.3109/07853 89010 9002089

Hestbaek, L., Leboeuf-Yde, C., & Manniche, C. (2003). Low back pain: What is the long-term course? A review of studies of general

patient populations. *European Spine Journal*, *12*(2), 149–165.

Juch, J. N., Maas, E. T., Ostelo, R. W., Groeneweg, J. G., Kallewaard, J.-W., Koes, B. W., … van Tulder, M. W. (2017). Effect of radiofrequency denervation on pain intensity among patients with chronic low back pain: The Mint randomized clinical trials. *JAMA*, *318*(1), 68–81. https ://doi.org/10.1001/jama.2017.7918

Klenerman, L., Slade, P., Stanley, I., et al. (1995). The prediction of chronicity in patients with an acute attack of low back pain in a general practice setting. *Spine*, *20*(4), 478–484. https ://doi.org/10.1097/00007 632-19950 2001-00012

Koopmanschap, M. A. (2005). PRODISQ: A modular questionnaire on productivity and disease for economic evaluation studies. *Expert Review of Pharmacoeconomics & Outcomes Research*, *5*(1), 23. https ://doi.org/10.1586/14737 167.5.1.23

Koopmanschap, M. A., & Rutten, F. F. (1996). A practical guide for calculating indirect costs of disease. *Pharmacoeconomics*, *10*(5), 460–466. https ://doi.org/10.2165/00019 053-19961 0050-00003

Lambeek, L. C., van Tulder, M. W., Swinkels, I. C., Koppes, L. L., Anema, J. R., & van Mechelen, W. (2011). The trend in total cost of back pain in The Netherlands in the period 2002 to 2007. *Spine*, *36*(13), 1050–1058. https ://doi.org/10.1097/BRS.0b013 e3181 e70488

Lamers, L., Stalmeier, P., McDonnell, J., & Krabbe, P. (2005). Measuring the quality of life in economic evaluations: The Dutch EQ-5D tariff. *Nederlands Tijdschrift Voor Geneeskunde*, *149*(28), 1574–1578.

Lancourt, J., & Kettelhut, M. (1992). Predicting return to work for lower back pain patients receiving worker's compensation. *Spine*, *17*(6), 629–640. https ://doi.org/10.1097/00007 632-19920 6000-00002

Linton, S. J., & Halldén, K. (1998). Can we screen for problematic

back pain? A screening questionnaire for predicting outcome in acute and subacute back pain. *Clinical Journal of Pain, 14*(3), 209–215. https://doi.org/10.1097/00002 508-19980 9000-00007

Linton, S. J., Nicholas, M., & Shaw, W. (2018). Why wait to address high-risk cases of acute low back pain? A comparison of stepped, stratified, and matched care. *Pain, 159*(12), 2437. https ://doi.org/10.1097/j.pain.00000 00000 001308

Lousberg, R., Van Breukelen, G. J., Groenman, N. H., Schmidt, A. J., Arntz, A., & Winter, F. A. (1999). Psychometric properties of the multidimensional pain inventory, dutch language version (MPIDLV). *Behavior Research and Therapy, 37*(2), 167–182. https ://doi.org/10.1016/S0005-7967(98)00137-5

Maas, E. T., Juch, J. N., Groeneweg, J. G., Ostelo, R. W. J. G., Koes, B. W., Verhagen, A. P., … van Tulder, M. W. (2012). Cost-effectiveness of minimal interventional procedures for chronic mechanical low back pain: Design of four randomised controlled trials with an economic evaluation. *BMC Musculoskeletal Disorders*, *13*(1), 260. https ://doi.org/10.1186/1471-2474-13-260

Maniadakis, N., & Gray, A. (2000). The economic burden of back pain in the UK. *Pain, 84*(1), 95–103. https ://doi.org/10.1016/S0304-3959(99)00187-6

McKillop, J. M., & Nielson, W. R. (2011). Improving the usefulness of the Multidimensional *Pain Inventory. Pain Research and Management, 16*(4), 239–244. https ://doi org/10.1155/2011/873424

Nordin, M., Hiebert, R., Pietrek, M., Alexander, M., Crane, M., & Lewis, S. (2002). Association of comorbidity and outcome in episodes of nonspecific low back pain in occupational populations. *Journal of Occupational and Environmental Medicine, 44*(7), 677–684. https ://doi.org/10.1097/00043 764-20020 7000-00015

Pincus, T., Burton, A. K., Vogel, S., & Field, A. P. (2002). A systematic review of psychological factors as predictors of

chronicity/disability in prospective cohorts of low back pain. *Spine, 27*(5), E109–E120. https ://doi.org/10.1097/00007 632-20020 3010-00017

Rabin, R. (2001). Charro Fd. EQ-SD: A measure of health status from the EuroQol Group. *Annals of Medicine, 33*(5), 337–343. https ://doi.org/10.3109/07853 89010 9002087

Ritzwoller, D. P., Crounse, L., Shetterly, S., & Rublee, D. (2006). The association of comorbidities, utilization and costs for patients identified with low back pain. *BMC Musculoskeletal Disorders. 7*(1), 72. https ://doi.org/10.1186/1471-2474-7-72

Skargren, E. I., & Öberg, B. E. (1998). Predictive factors for 1-year outcome of low-back and neck pain in patients treated in primary care: Comparison between the treatment strategies chiropractic and physiotherapy. *Pain, 77*(2), 201–207. https ://doi.org/10.1016/S0304-3959(98)00101-8

Steyerberg, E. W. (2010). *Clinical prediction models. A practical approach to development, validation and updating.* New York, NY: Springer.

Steyerberg, E. W., Moons, K. G., van der Windt, D. A., Hayden, J. A., Perel, P., & Schroter, S. (2013). Prognosis research strategy (PROGRESS) 3: Prognostic model research. *PLoS Medicine. 10*(2), e1001381. https ://doi.org/10.1371/journ al.pmed.1001381

Steyerberg, E. W., Vickers, A. J., Cook, N. R., Gerds, T., Gonen, M., Obuchowski, N., … Kattan, M. W. (2010). Assessing the performance of prediction models: A framework for some traditional and novel measures. *Epidemiology (Cambridge, Mass), 21*(1), 128. https://doi.org/10.1097/EDE.0b013 e3181 c30fb2

System TCFCoaHPH. (2012). *The performance improvement imperative: Utilizing a coordinated, community-based approach to enhance care and lower costs for chronically Ill patients.*

Tsuboi, Y., Murata, S., Naruse, F., & Ono, R. (2019). Association between pain-related fear and presenteeism among eldercare

workers with low back pain. *European Journal of Pain, 23*(3), 495–502. https ://doi.org/10.1002/ejp.1323.

Valat, J.-P., Goupille, P., & Védere, V. (1997). Low back pain: Risk factors for chronicity. *Revue Du Rhumatisme. English Edition, 64*(3), 189–194.

van Tulder, M., Koes, B., & Bombardier, C. (2002). Low back pain. *Best Practice & Research Clinical Rheumatology, 16*(5), 761–775. https://doi.org/10.1053/berh.2002.0267

Vander Zee, K. I., Sanderman, R., Heyink, J. W., & Haes, H. (1996). Psychometric qualities of the RAND 36-Item Health Survey 1.0: A multidimensional measure of general health status. *International Journal of Behavioral Medicine, 3*(2), 104. https ://doi.org/10.1207/s1532 7558i jbm03 02_2

Verkerk, K., Luijsterburg, P. A., Heymans, M. W., Ronchetti, I., Pool-Goudzwaard, A. L., Miedema, H. S., & Koes, B. W. (2013). Prognosis and course of disability in patients with chronic nonspecific low back pain: A 5-and 12-month follow-up cohort study. *Physical Therapy, 93*(12), 1603–1614. https ://doi.org/10.2522/ptj.20130076

Vlaeyen, J. W. S., Maher, C. G., Wiech, K., van Zundert, J., Meloto, C. B., Diatchenko L., … Linton, S. J. (2018). Low back pain. *Nature Reviews Disease Primers, 4*(1), 52. https ://doi.org/10.1038/s41572-018-0052-1

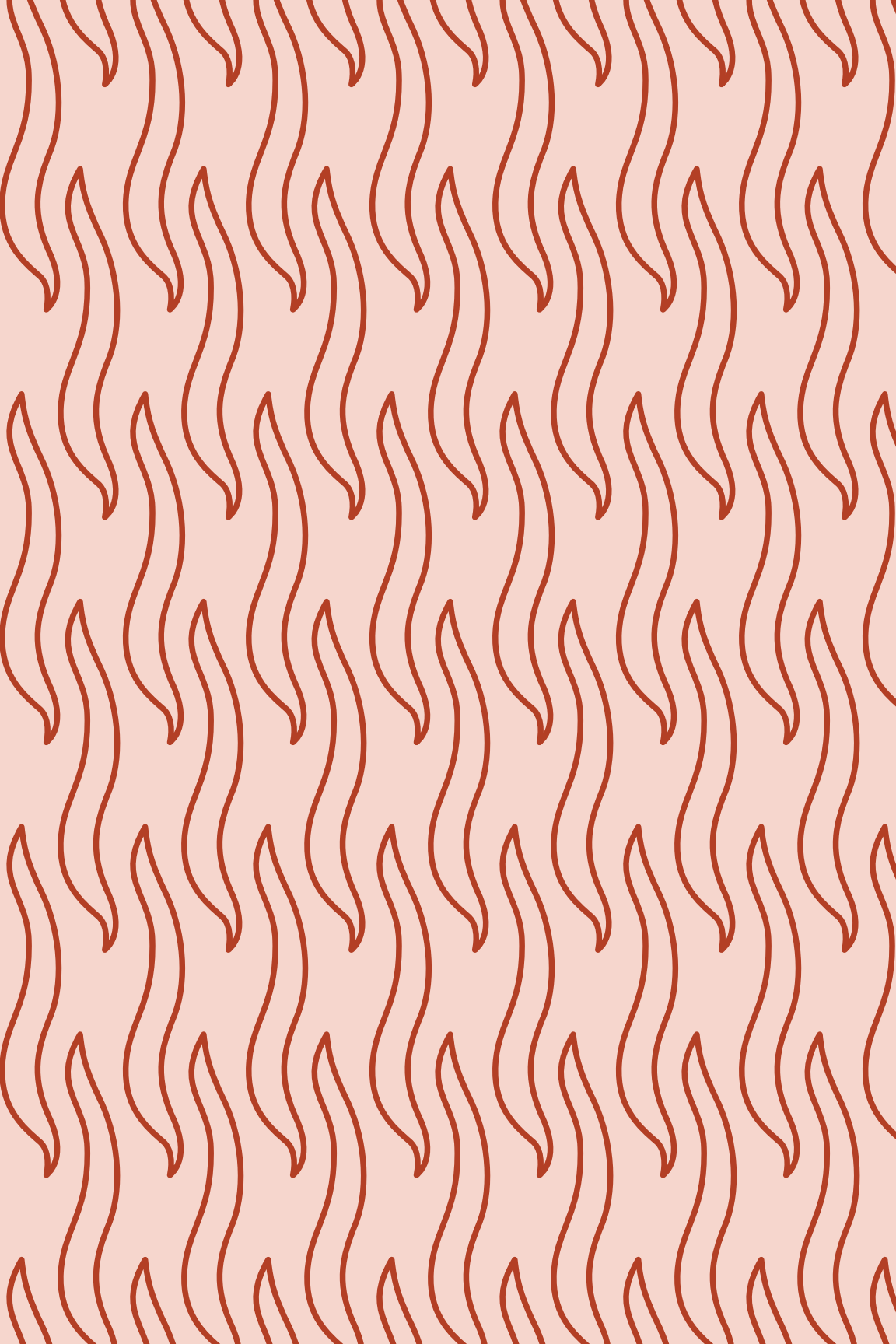
Vos, T., Abajobir, A. A., Abate, K. H., Abbafati, C., Abbas, K. M., Abd-Allah, F., … Murray, C. J. L. (2017). Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: A systematic analysis for the Global Burden of Disease Study 2016. *The Lancet, 390*(10100), 1211–1259. https ://doi.org/10.1016/S0140-6736(17)32154-2

Waddell, G. (2006). Preventing incapacity in people with musculoskeletal disorders. *British Medical Bulletin, 77*(1), 55–69. https ://doi.org/10.1093/bmb/ldl008

Walker, B., Muller, R., & Grant, W. (2003). Low back pain in Australian adults: The economic burden. *Asia-Pacific Journal of Public Health*, *15*(2), 79–87. https ://doi.org/10.1177/10105 39503 01500202

Wenig, C. M., Schmidt, C. O., Kohlmann, T., & Schweikert, B. (2009). Costs of back pain in Germany. *European Journal of Pain*, *13*(3), 280–286. https ://doi.org/10.1016/j.ejpain.2008.04.005

Wieser, S., Horisberger, B., Schmidhauser, S., Eisenring, C., Brügger, U., Ruckstuhl, A., … Müller, U. (2011). Cost of low back pain in Switzerland in 2005. *The European Journal of Health Economics*, *12*(5), 455–467. https ://doi.org/10.1007/s10198-010-0258-y



B

Effectiveness and cost-effectiveness of sciatica treatments

4

Cost-effectiveness of combination therapy (Mechanical Diagnosis and Treatment and Transforaminal Epidural Steroid Injections) among patients with an indication for a Lumbar Herniated Disc surgery: Protocol of a randomized controlled trial

Mutubuki, E.N.,[1,5] van Helvoirt, H.,[2] van Dongen, J.M.,[1,5] Vleggeert-Lankamp,[3] C.L.A., Huygen, F.J.P.M.,[4] van Tulder, M.W.,[1,6] Klopper-Kes, A.H.J.,[2] Ostelo, R.W.J.G.[1,5]

[1] Department of Health Sciences, Faculty of Science, Vrije University Amsterdam, Amsterdam Movement Sciences, The Netherlands
[2] Stichting Rugpoli, Delden, The Netherlands
[3] Department of Neurosurgery, Leiden University Medical Centre, Leiden, The Netherlands
[4] Department of Anesthesiology, Centre of Pain Medicine, Erasmus MC, Rotterdam,
[5] Department of Epidemiology and Biostatistics, VU University Medical Center, Amsterdam Movement Sciences, The Netherlands
[6] Department of Physiotherapy & Occupational Therapy, Aarhus University Hospital, Aarhus, Denmark

Mutubuki EN, van Helvoirt H, van Dongen JM, et al. Cost-effectiveness of combination therapy (Mechanical Diagnosis and Treatment and Transforaminal Epidural Steroid Injections) among patients with an indication for a Lumbar Herniated Disc surgery: Protocol of a randomized controlled trial. Physiother Res Int. 2020;25(1):e1796. doi:10.1002/pri.1796

ABSTRACT

Objectives: The general consensus is that surgical treatment is advised when conservative methods fail in patients with lumbosacral radicular syndrome (LRS). Preliminary evidence from our pilot study indicates that combination therapy (mechanical diagnosis therapy and transforaminal epidural injections) can prevent surgical treatment in patients on the waiting list for surgery. The pilot study lacked a control group and therefore firm conclusions pertaining to effects could not be made. This study aims to determine if combination therapy, performed while being on the waiting list for lumbar herniated disc surgery, is effective and cost-effective compared to usual care (i.e. no intervention while being on the waiting list) among patients with an MRI-confirmed indication for a lumbar herniated disc surgery.

Methods: A randomised controlled trial will be conducted with an economic evaluation. Patients aged 18 and above with incapacitating LRS, with leg pain and an MRI confirmed indication for lumbar disc hernia surgery, will be recruited from 7 Dutch hospitals. While being on the waiting list for lumbar herniated disc surgery, patients will be randomised to either the combination therapy or usual care group. The primary outcome measure is the number of patients undergoing lumbar disc surgery during 12-months follow-up. Secondary outcomes include back and leg pain intensity (NPRS), physical functioning (RMDQ-23), self-perceived recovery (GPE), and health-related quality of life (EQ-5D-5L and SF12). For the economic evaluation, societal and healthcare costs will be measured. Measurements moments are baseline, 1, 2, 4, 6, 9 and 12 months. Data will be analysed according to the intention-to-treat principle.

Conclusion: No randomized controlled trials have evaluated the effectiveness and cost-effectiveness of combination therapy compared to usual care in patients with an indication for lumbar herniated disc surgery, which emphasizes the importance of this study.

Keywords: epidural injections, physiotherapy, radiculopathy

INTRODUCTION

Lumbar disc herniation is the most common cause of lumbosacral radicular syndrome (LRS); also known as sciatica. Characteristics of LRS include radiating lower limb pain into a particular dermatome which may be accompanied by sensory and or motor deficits (Oosterhuis et al., 2014). Evidence suggests that the pathophysiology of LRS is not attributed to just pressure on the nerve roots but to a complex interplay of inflammatory, immunological and pressure related processes (Stafford, Peng, & Hill, 2007). Estimated LRS incidence rate in Western countries is 5 per 1000 (Cherkin, Deyo, Loeser, Bush, & Waddell,1994). In the Netherlands, the incidence rate of LRS in general practice is 12 per 1,000 patients per year (Schaafstra et al., 2015). The yearly direct and indirect costs of LRS are high and estimated to be 1.2 billion euros in the Netherlands (Health Council of the Netherlands 1999).

There is a lot of variation in LRS prevalence in literature (Konstantinou, & Dunn, 2008). Consequently, there are disparities in spinal surgery rates regionally and internationally (Weinstein, Lurie, Olson, Bronner, & Fisher, 2006). In the United States, spinal surgery rates are 30% higher than in the Netherlands, 80% higher than in UK and 50-60% higher than in Canada (Cherkin et al., 1994).

The Dutch guideline 'Lumbosacral Radicular Syndrome' recommends surgical treatment if the radiating leg pain persists following conservative management (CBO, 2008). A significant number of patients undergoing surgery for lumbar disc herniation suffer residual complaints. Recovery rates in the literature vary wildly. Recent figures from the Netherlands suggest a rate between 69%-79% after two-year follow-up and 10%-15% of the patients need repeated surgery, the majority of which were due to recurrent disc herniation at the same level (Arts et al., 2011). Findings of Peul et al., (2007) indicate that surgical and non-surgical management of lumbar hernia are equally successful in the long-term.

Both mechanical diagnosis and treatment (MDT) and transforaminal epidural steroid injections (TESIs) are reported to be individually effective in reducing pain and improving function

among LRS patients (Chou et al., 2015; van Helvoirt et al., 2014). Epidural corticosteroid injections for radiculopathy are associated with immediate reduction in pain (Chou et al., 2015). TESIs is indicated in LRS and the role of physiotherapy, possibly in combination with TESIs should be further explored (CBO 2008). None of the RCTs included by Chou et al., (2015) combined TESIs with MDT. The only two publications that we identified assessing a combination therapy of TESIs and MDT were our own pilot study (van Helvoirt et al., 2014), and a report on three cases of acute cervical radiculopathy (Desai, Padmanabhan, Simbasivan, Kamanga-Sollo, & Dharmappa, 2012). Our pilot study suggests that combining these interventions has the potential to reduce the number of lumbar herniated disc surgeries, as only 22% of patients with a herniated lumbar disc still needed surgery after one-year follow-up (van Helvoirt et al., 2014).

Research indicates that, the effects of lumbar disc surgery are comparable to those of conservative treatment after one and two years (Peul et al, 2007; Peul et al. 2008). Clinical guidelines prescribe shared decision making and that pros and cons of both options should be discussed with patients. (CBO, 2008; NICE, 2016). Surgery is costly and potentially causes various side effects (e.g. nerve root damage, infection, pain that continues after surgery), hence spinal surgeons typically aim to prolong conservative therapy. Physiotherapists could play an important role in preventing surgery if they combine their treatment with optimal pain management (Schaafstra et al., 2015). Although the results from our pilot study seem to be promising, the effectiveness and cost-effectiveness of a combination therapy of TESIs and MDT has not been rigorously evaluated. Therefore, this study aims to determine if a combination therapy, while being on the waiting list for a lumbar herniated disc surgery, is effective and cost-effective compared to usual care (i.e. no intervention while being on the waiting list) among patients with an MRI-confirmed indication for a lumbar herniated disc surgery.

METHODS

Study design:

A multicentre randomized controlled trial with a 12-month follow-up and a full economic evaluation.

Ethical approval
In September 2017, the Medical Ethics Committee of the VU University Medical Centre Amsterdam approved the study protocol, registration number NL60558.029.17 and the study was registered in the Dutch Trial Register NTR6715.

Subjects:

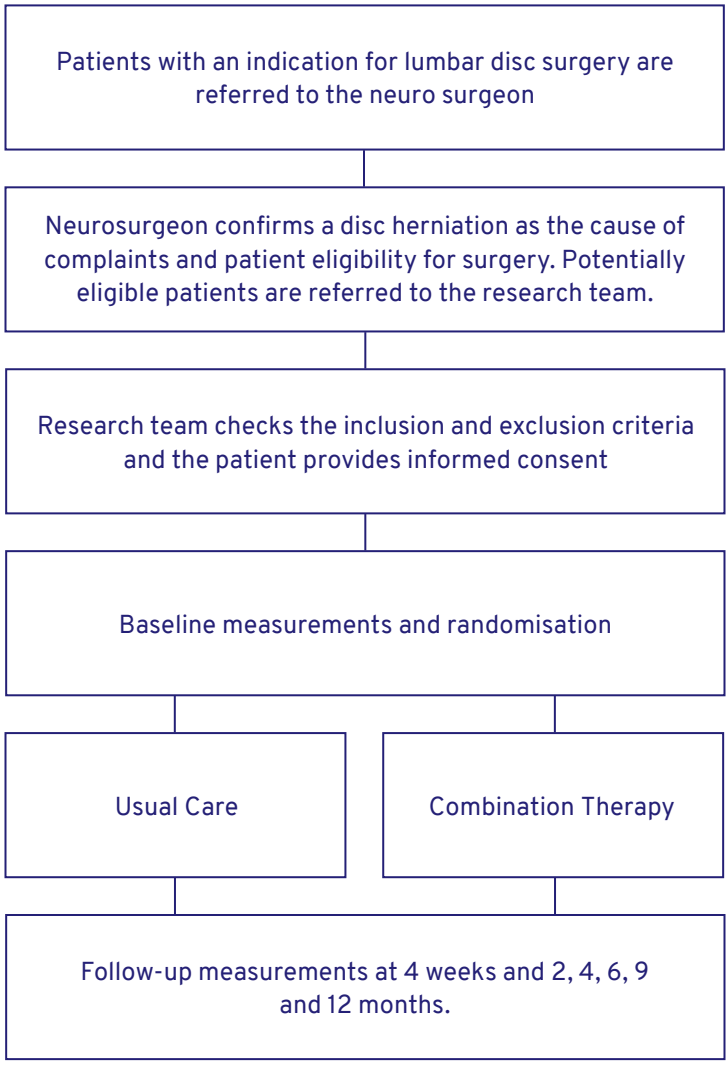
Inclusion criteria
Patients aged 18 and above; eligibility for lumbar disc hernia surgery; incapacitating LRS with leg pain (numeric pain rating scale (NPRS)>6; with or without back pain) that had lasted for a minimum of 6 weeks with or without mild neurological deficit (i.e. Medical Research Council [MRC]>3); an MRI which confirms a hernia nuclei pulposi (HNP) that compromises the spinal nerve and can explain the clinical symptoms of the patient.

Exclusion criteria
Spine surgery and or transforaminal injections at the same level during the previous 6 months; bony stenosis; cauda equina syndrome; spondylolisthesis; pregnancy; complicated disc herniation requiring more than one operation; severe coexisting disease (e.g. osteoporosis, dementia); patient with contra-indications for steroid injections; insufficient knowledge of the Dutch language; emergency surgery as determined by the neurosurgeon; and being allergic for Iohexol 240mg/ml (i.e. OMNIPAQUE 240).

Patient recruitment
Figure 1 shows the flow diagram of the study. Neurosurgeons or orthopaedic surgeons of the participating hospitals recruit and inform the patients about the study and the possibility to

participate while they are on the waiting list for surgery. The surgeon then refers the patient to the research team who will check if the patient meets the aforementioned inclusion and exclusion criteria. If the patient is eligible and gives informed consent, the patient will be included in this study. Baseline measures will then be made and patients will be randomized to either the intervention or the control arm of the study.

Figure 1: Flow chart of the PLUS study



Setting

Participants will be recruited from 7 hospitals in the Netherlands. The hospitals were chosen due to their proximity to the four primary care based outpatient clinics (so-called 'Rugpoli's') where the combination therapy will be provided.

Materials:

Prognostic factors

Prognostic factors measured at baseline using an online questionnaire include, duration and severity of complaints before operation, various psychosocial variables (somatization, distress, anxiety), and known confounding factors such as age, gender, educational level and treatment expectation (CBO, 2008).

Primary outcomes measure

The primary outcome measure is the proportion of LRS patients undergoing lumbar disc surgery during 12-month follow-up. Patients will be scored as either having had a lumbar surgery or not. For this purpose, patients will be asked whether they had a lumbar surgery during the previous weeks following the last assessment, using an online questionnaire at 2, 4, 6, 9 and 12 months after randomisation, and validated using hospital records.

Secondary outcomes measures

In line with the core outcome set for clinical research and clinical practice (COS) (Chiarotto et al. 2018), secondary study parameters include back and leg pain, physical functioning, and health-related quality of life. Additionally we will measure self-perceived recovery, patient satisfaction (single question) and pain location (pain mannequin). Measurements will take place at baseline, 2, 4, 6, 9 and 12 months after randomization and administered through online questionnaires. An additional pain intensity measurement will be carried out at 4 weeks after randomisation. The rationale is that, especially leg pain will improve in the combination group. Complications will be noted in a case report form (CRF). Societal and healthcare costs will be estimated for economic evaluation using resource use data, collected through online questionnaires at baseline, 2, 4, 6, 9

and 12 months after randomisation.

Back and leg pain will be measured using the (Numeric Pain Rating Scale [NPRS])
The NPRS ranges from 0 ("no pain") to 10 ("representing worst pain imaginable"). The numerical pain rating scale is reliable, valid and has good sensitivity (Williamson, & Hoggart, 2005).

Functional status will be measured with (Roland Morris Disability Questionnaires [RMDQ-23])
5 RMDQ 24 items were removed and 4 new items were added from the initial source of RMDQ items, to create RMDQ 23 (Bergner, Bobbitt, Carter, & Gilson, 1981). RMDQ 23 consists of 23 "yes" or "no" questions, measuring limitation in activity associated with back and leg pain (Kent, Grotle, Dunn, Albert, & Lauridsen, 2015). The scores can range from 0 (no disability) to 23 (maximal disability). It has been extensively used in radiculopathy and stenosis research as a standardized measure and is widely used to assess disability specific to back and leg pain making it suitable people with LRS (Kent et al., 2015). The RMDQ-23 is regarded as reliable and valid (Yamato, Maher, Saragiotto, Catley, & McAuley, 2017). It has been translated into Dutch.

Health related quality of life (EQ-5D-5L and SF-12)
The EQ-5D-5L is a quality of life scale that is responsive for chronic low back pain patients (Soer, Reneman, Speijer, Coppes, & Vroomen, 2012). The EQ-5D-5L has five health dimensions: mobility, self-care, daily activities, pain/discomfort, and anxiety/depression. Within each dimension, the patients can self-rate their level of severity; no, slight, moderate, severe problems, unable to perform or do the task (Versteegh et al., 2016). For the economic evaluation, the patients' EQ-5D-5L health states will be converted into utility scores ranging from 0 ("death") to 1 ("optimal health") using the Dutch tariff (Versteegh et al., 2016).

The SF-12 is a shorter version of the SF-36 health-related quality-of-life questionnaire. The SF-12 has been proven to be a reliable and valid questionnaire for low back pain (Xuemei et at., 2003). The questionnaire relates to the analysis of the general functional status of patients. It consists of 12 questions from the following 8 domains; 1) physical functioning, 2) physical role

limitations, 3) emotional role imitations, 4) social functioning, 5) physical pain, 6) general mental health, 7) vitality, and 8) general health perception. These eight domains can be summarized into a physical and psychological main domain (Xuemei et at., 2003). For the economic evaluation, Quality Adjusted Life Years (QALYs) will be calculated by multiplying the patients' time spent in a certain health state by the respective utility value (i.e. area under the curve method).

Self-perceived recovery will be measured using the Global Perceived Effect (GPE):
The GPE measures a patient's self-perceived recovery using a 7-point scale ranging from "worse than ever" (1) to "completely recovered" (7). Being recovered will be defined as being "completely recovered" or "much improved"; other responses will be defined as not recovered. The test reliability of the GPE scale is said to be good (Kampera et al. 2010).

Societal as well as healthcare costs
Societal costs include costs of the intervention, other healthcare use, informal care, unpaid productivity losses, and costs due to absenteeism (i.e. sickness absence) and presenteeism (i.e. being less productive while being at work). Health care costs will only include costs accruing to the formal Dutch healthcare sector.

Intervention costs will be micro-costed. For this purpose, information about the combination therapy will be gathered using a case report form (CRF), including information on patient classification, number of sessions, discharge and referral to a network MDT or to the surgeon. All other cost categories will be measured using cost questionnaires administered at baseline 2, 4, 6, 9 and 12 months after randomisation. Resource use will be valued in accordance with the Dutch manual of costing (Hakkaart-van Roijen, Tan, & Bouwmans, 2010). See Table 1: overview of the data collection.

Table 1: Overview of the data collection

Outcome measures	Baseline	4 weeks	2 months	4 months	6 months	9 months	12 months
Prognostic factors							
Patient demographics and prognostic factors	X						
Primary outcome							
Surgery rate			X	X	X	X	X
Secondary outcomes							
Pain leg (NPRS)	X	X	X	X	X	X	X
Pain back (NPRS)	X	X	X	X	X	X	X
Health-related Quality of Life (SF-12, EQ-5D-5L)	X		X	X	X	X	X
RMDQ-23	X		X	X	X	X	X
Self-perceived recovery (GPE)			X	X	X	X	X
Societal and healthcare costs (Cost questionnaires)	X		X	X	X	X	X
Patient satisfaction, complications, pain location	X		X	X	X	X	X

Procedure:

Treatment allocation

Randomization will be done by an independent researcher who is not involved in treatment procedures, using a web-based randomization program. The randomization sequence is developed centrally. Therefore, the independent researcher does not have any influence on the randomization procedure and the treatment allocation is concealed. Patients will be randomized at the individual level and in a 1:1 ratio. We will stratify on duration of complaints (i.e. <6 months versus ≥6 months), and we will use one randomization list per hospital. The randomization key will be safeguarded by an independent researcher. Patients allocated to the intervention group will be called for an appointment within 48 hours of randomization and will attend their first appointment within the first week following randomization.

Combination therapy intervention

While being on the waiting list to receive lumbar disc surgery, intervention group participants will receive the combination therapy. The combination therapy has two parts: 1) Mechanical Diagnosis and Treatment [MDT]; 2) Transforaminal Epidural Steroid Injections [TESIs] and is delivered by teams of pain interventionists and physiotherapists. The pain interventionist is responsible for the TESIs and the physiotherapists for the MDT. Prior to receiving the combination therapy, patients are seen by a pain interventionist who checks for contraindications for injections and medications including steroid use. During the same appointment, participants are classified as "centralizers" or "non-centralizers" using MDT principles, that is, assessment of the patients' pain pattern responses on repeated movement tests.

Centralizers

Centralizers are defined as patients with centralization (i.e. a clear change in leg pain location from a more peripheral location towards a more central location, which lasts after testing staying in neutral) or directional preference (i.e. a reduction in pain intensity, but not in location, which lasts after testing staying in neutral). Testing for centralization is done according to MDT

principles as described in the textbook of McKenzie (2003). Searching for centralization is done during repeated movement testing or sustained positioning in a certain direction. This direction differs in patients. Centralization could be found in extension, flexion, side gliding, rotation or a combination. While testing, an MDT trained physiotherapist is able to decide how many repetitions are needed (usually between 10 to 20) and can add manual force (therapist overpressure or mobilization) if needed, depending on pain response, during and after testing. The MDT system appears to have acceptable inter-rater reliability for classifying patients with back pain into main/sub-syndromes, when applied by therapists who have completed the credentialing examination, but unacceptable reliability in other therapists (Garcia, et al. 2018). Sustained positioning is often used with high levels of leg pain and major movement loss in range of motion (ROM) testing. If a certain direction is found to centralize the leg and back pain, that same direction of exercise or positioning will be used as the initial treatment direction. Patients are advised to exercise 7 to 8 times a day (i.e. 10 repetitions of extension in lying 8 times per day). In the process of centralization, patients get postural advice in the direction of centralization (i.e. if extension is the centralizing direction, patients will be advised to keep their back in lordosis and avoid flexion movements for a period of time). As soon as centralization is full and stable, the physiotherapist trained in MDT will restore full function and most importantly check for fear avoidance of the restricted movement direction, as this could be the case after avoiding a certain direction for a period of time.

Non centralizers

Non-centralizers are defined as patients with peripheralization (i.e. a clear change towards a more peripheral leg pain location or an increase in leg pain) or no effect (i.e. no change in leg pain location or intensity). The hypothesis is that in these patients the pain is a result of the inflammation. Non-centralizers will receive a transforaminal epidural injection in accordance with the procedure described below:

Procedure for the Transforaminal Epidural Steroid Injections [TESIs]

The patient lies in prone position. Under fluoroscopic guidance with contrast medium (Iohexol 240mg/ml 0,5cc), a very thin needle will be placed next to the compressed nerve. The contrast medium is then used to control if fluid will come to the compressed nerve. After which a combination of a local anaesthetic (lidocaine 20mg/ml 0.5cc) with an anti-inflammatory agent (dexamethasone 20mg/ml 0,5cc) is injected. Half an hour after the injection, the pain interventionist checks the effects of the injection. The duration of pain absence is dependent on the working of the anti-inflammatory drug on the inflammation. Two weeks after the injection, the patient is seen back by the MDT therapist to check for classification in the described subgroups and to decide if a second injection is necessary in shared decision making with patient and the pain interventionist. If pain reduction is less than 80%, then usually a second injection will be administered with patient consent. A maximum of three injections are given to optimize pain relief.

Following Transforaminal Epidural Steroid Injections [TESIs]
After the injections, participants will be re-classified using the same MDT principles, into 4 subgroups. The sub-groups are as follows;

1) Resolved symptoms (i.e. no or irrelevant pain; ≤1 on a 0-10 NPRS),
2) Centralizing and significantly less pain (i.e. a pain reduction of ≥2 on a 0-10 NPRS),
3) Non-centralizing and significantly less pain (i.e. a pain reduction of ≥2 on a 0-10 NPRS),
4) Non-centralizing with high levels of pain and disability (i.e. a pain score of ≥8 on a 0-10 NPRS and a disability score of >10 on the RDMQ-23).

In the first three sub-groups mentioned here, patients will get specific MDT exercises and advice. These 3 sub-groups will be treated by an MDT therapist in 1 to 6 sessions in on average 4 weeks. If more sessions are required, patients are referred to accredited MDT therapists within the Rugpoli network. Network therapists are in close contact with the Rugpoli centres and are located all over the Netherlands.

Only the patients in subgroup 4 will be referred back to the neurosurgeon who will assess whether patients still require

surgery. Throughout the combination therapy, there is shared decision-making.

Control intervention

Control group participants will solely be placed on a waiting list and scheduled to receive lumbar disc surgery if still required. The aim of surgery is to remove the symptomatic disc herniation by a minimal unilateral transflaval approach with magnification, with the patient under general or spinal anaesthesia.

Use of co-intervention

Use of co-interventions by patients is allowed and will be monitored. Patients will be requested to complete questionnaires in which medication usage and any healthcare utilization is recorded throughout the follow-up period.

Blinding

We will not attempt to blind the patients to the intervention or control condition, as this is practically impossible in this study due to the nature of the intervention. The outcomes assessor will not be blinded, because all outcomes are self-reported. Treatment providers (i.e. surgeons, physiotherapist and anaesthesiologists) will not be blinded due the nature of the intervention they will provide to patients.

Sample size

We expect that 90-95% (93% was used in the sample size calculation) of the patients in the usual care group will receive surgery and we hypothesize that in the combination group this rate will be reduced by 30% (or more). To detect this difference of 30% with an alpha of 0.05 (two-sided), a power of 95%, anticipating a 20% drop-out rate, and taking into account the multilevel structure (with an ICC of 0.15), we need to include a total of 146 patients (n=73 per treatment group) (Pocock, 1983; Twisk, 2006). Even though participants will not be randomized at the hospital-level, the multilevel structure of the data was accounted for in the sample size calculation, because patients are recruited from different hospitals (i.e. patients recruited from one hospital are likely to be more similar than those recruited from other hospitals) and clusters will likely not be balanced.

Statistical/data analysis

Baseline characteristics of the patients in both study groups will be presented using descriptive statistics (mean [standard deviation], median [range] or proportion) to assess if balanced groups were obtained after randomization (i.e. having an equal distribution of the main outcome measures, prognostic factors and known confounding factors such as age, gender, educational level and treatment expectation).

Primary outcome analysis

The primary analysis will be an intention-to-treat analysis. The primary study parameter (i.e. surgery; yes/no) will be analysed in a logistic mixed model with responses at 2, 4, 6, 9 and 12 months. In this analysis, we will take into account the levels of hospital, patient and time of measurement. An odds ratio with 95%CI between the combination therapy group and usual care group will be calculated. If necessary, the analysis will be adjusted for important prognostic characteristics.

Secondary outcome analysis

The secondary study parameters, back pain, leg pain, self-perceived recovery, health-related quality of life and functional status, will be analysed in the same way as the primary study parameter (i.e. surgery). However, for continuous outcomes, we will use a linear mixed model with the same multilevel structure.

Economic evaluation

An economic evaluation will be performed from a societal and a healthcare perspective. When the societal perspective is applied, all costs and consequences relevant to the intervention will be taken into account irrespective of whom pays or benefits, whereas only those accruing to the formal Dutch healthcare sector will be considered when the healthcare perspective is applied (Hakkaart-van Roijen, van der Linden, Bouwmans & Kanters, 2015).; Brouwer, van Exel, Baltussen, & Rutten, 2006).

The economic evaluation will be performed in accordance with the intention-to-treat principle and in terms of QALYs as well as the primary outcome proportion of LRS patients undergoing lumbar disc surgery during 12-month follow-up. Missing data will

be imputed using multiple imputations by chained equations. The imputation model will be constructed in accordance with the recommendations of (White, Royston, & Wood, et al, 2011). Imputed datasets will be analysed as specified below, after which pooled estimates will be estimated using Rubin's rules (White, Royston, & Wood, 2011).Incremental cost-effectiveness ratios will be calculated by dividing the difference in costs by that in effects. In order to account for the possible clustering of data, analyses will be performed using linear mixed models (Gomes, 2012). Accounting for the possible clustering of data (e.g. at the hospital level) is very important, as most economic evaluations fail to do so, whereas ignoring the possible clustering of data might lead to inaccurate levels of uncertainty and inaccurate point estimates (Gomes, 2012). Bootstrapping techniques will be used to estimate the uncertainty surrounding the cost-effectiveness estimates. Uncertainty will be shown by plotting cost-effect pairs on cost-effectiveness planes and by constructing cost-effectiveness acceptability curves (Black, 1990; Drummond, Sculpher, Torrance, O'Brien, Stoddart, 2005; Fenwick, O'Brien, Briggs, 2004). Cost-effectiveness acceptability curves indicate the probability of an intervention being cost-effective compared with a control for a range of ceiling ratios (i.e. the maximum amount of money decision-makers are willing to pay per unit of effect gained).

Various one-way sensitivity analyses will be performed to test the robustness of the study results (e.g. complete-case analysis, per-protocol analysis).

DISCUSSION

Prior to this study, only our pilot study has evaluated the effectiveness of a combination therapy (MDT and TESIs) in patients presenting with lumbar disc herniation (van Helvoirt et al., 2014). The pilot study indicated the importance of a multidisciplinary approach which addresses the inflammatory and mechanical contributors to spine mediated pain, and has the potential to reduce the numbers of herniated disc surgery. However the cost-effectiveness of the combination therapy was

not explored. The results of the pilot study should be interpreted with great caution because of lack of a control group.

The importance of this study is further emphasized by the fact that there are huge discrepancies in the treatment and management of lumbar disc herniation regionally and internationally (Weinstein et al., 2006). In addition, a lot of costs and burden to society have been associated with LRS (Health Council of the Netherlands, 1999). Therefore, the combination therapy may not only benefit individual patients, but also society as a whole.

To our knowledge this is the first randomized controlled trial investigating the effectiveness and cost-effectiveness of the combination therapy in patients with LRS compared to usual care. Therefore, the present study aims to determine if a combination therapy, while being on the waiting list for a lumbar herniated disc surgery, is effective and cost-effective compared to usual care among patients with an indication for a lumbar herniated disc surgery. Hence, this research will help bridge the knowledge gap in the treatment and management of patients with LRS.

REFERENCES

Arts, M.P., Brand, R., van den Akker, M.E., Koes, B.W., Bartels, R.H., Tan, W.F., & Peul, W.C. (2011). Tubular Diskectomy vs Conventional Microdiskectomy for the Treatment of Lumbar Disk Herniation: 2-Year Results of a Double-Blind Randomized Controlled Trial. *Neurosurgery, 69 (1),135–144.*

Bergner, M., Bobbitt, R.A., Carter, W.B., Gilson, B.S. (1981). The Sickness Impact Profile: Development and Final Revision of a Health Status Measure. *Medical Care, 19(8), 787-805.*

Black, W.C. (1990). The CE plane: a graphic representation of cost-effectiveness. *Medical Decision Making,10(3),212-214.*

Brouwer, W.B., van Exel, N., Baltussen, R.M., Rutten, F.F. (2006). A dollar is a dollar is a dollar--or is it? *Value in Health, 9(5),341-347.*

Kwaliteitsinstituut voor de Gezondheidszorg CBO (2008). *Richtlijn Lumbosacraal Radiculair Syndroom.* Utrecht

Cherkin, D.C., Deyo, R.A., Loeser, J.D., Bush, T., Waddell, G. (1994). An international comparison of back surgery rates. *Spine,19,1201-1206.*

Chiarotto A, Boers M, Deyo RA, Buchbinder R, Corbin TP, Costa LOP, Foster NE, Grotle M, Koes BW, Kovacs FM, Lin CC, Maher CG, Pearson AM, Peul WC, Schoene ML, Turk DC, van Tulder MW, Terwee CB, Ostelo RW. (2018). Core outcome measurement instruments for clinical trials in nonspecific low back pain. *Pain. 2018;159(3):481-495.*

Chou, R., Hashimoto, R., Friedly, J., Fu, R., Bougatsos, C., Sullivan, S.D., Jarvik, J. (2015). Epidural Corticosteroid Injections for Radiculopathy and Spinal Stenosis: A Systematic Review and Meta-analysis. *Ann Intern Med.,163(5),373-381.*

Desai, M.J., Padmanabhan, G., Simbasivan, A., Kamanga-Sollo,

G.G., Dharmappa, A. (2012). Directional Preference Following Epidural Steroid Injection in Three Patients with Acute Cervical Radiculopathy. *Pain Practice, 13(7), 559–565.*
Drummond, M.F., Sculpher, M.J., Torrance, G.W., O'Brien, B.J., Stoddart, G.L. (2005). *Methods for the Economic Evaluation of Health Care Programmes* (3rd ed.). New York: Oxford University Press.

Fenwick, E., O'Brien, B.J., Briggs, A. (2004). Cost-effectiveness acceptability curves-facts fallacies and frequently asked questions. *Health Economics. 2004; 13(5):405-415.*

Garcia, A.N., da Cunha Menezes Costa, L., de Souza, F.S., de Almeida, M.O., Araujo, A.C., Hancock, M., Costa, L.O.P. (2018). Reliability of Mechanical Diagnosis and Therapy system in patients with spinal pain: A systematic review. *JOSPT*

Gomes M., Grieve, R., Nixon, R., Ng, E.S., Carpenter, J., Thompson, S.G. (2012). Methods for covariate adjustments in cost-effectiveness analysis that use cluster randomised trials. *Health Economics, 21(9),1101-1118.*

Hakkaart-van Roijen, L., Tan, S.S., Bouwmans, C.A. (2010). *Handleiding Voor Kostenonderzoek.* College Voor Zorgverzekeringen.

Hakkaart-van Roijen, L., van der Linden, N., Bouwmans, C., Kanters, T., Tan, S.S., (2015). Kostenhandleiding: Methodologie van kostenonderzoek en referentieprijzen voor economische evaluaties in de gezondheidszorg.

Health Council of the Netherlands (1999). *Management of the lumbosacral radicular syndrome (sciatica).* The Hague: Health Council of the Netherlands, *publication no.1999/18.* ISBN 90-5549-284-1

Kampera, S.J., Ostelo, R.W., Knol, D.L., Maher, C.G., de Vet, H.C., Hancock, M.J. (2010). Global Perceived Effect scales provided reliable assessments of health transition in people with musculoskeletal disorders, but ratings are strongly influenced by

current status. *Journal of Clinical Epidemiology, 63(7), 760-766.*

Kent, P., Grotle, M., Dunn, K.M., Albert, H.B., Lauridsen, H.H. (2015). Rasch Analysis of the 23-item version of the Ronald Morris disability questionnaire. *Journal of Rehabilitation Medicine, 47(4), 356-364.*

Konstantinou, K., Dunn, K.M. (2008). Sciatica Review of epidemiological studies and prevalence estimates. *Spine, 33(22), 2464-2472.*

McKenzie, R., May, S. (2003). The Lumbar Spine. Mechanical Diagnosis and Therapy. Vol. II: New Zealand: *Waikanae Spinal publications Ltd.,731 s.* ISBN 0-9583647-6-1

National Institute for Health and Clinical Excellence (2016). Low back pain and sciatica in over 16s: assessment and management. *NICE guideline (NG59).*

Oosterhuis, T., Costa, L.O.P., Maher, C.G., de Vet, H.C., van Tulder, M.W., Ostelo, R.W. (2014). Rehabilitation after lumbar disc surgery. *Cochrane Database of Systematic Reviews, 3.*

Peul, W.C., van Houwelingen, H.C., van den Hout, W.B., Brand, R., Eekhof, J.A., Trans, J.T., Thomeer, R.T., Koes, B.W., Leiden-The Hague Spine Intervention Prognostic Study Group (2007). Surgery versus Prolonged Conservative Treatment for Sciatica. *The New England Journal of Medicine, 356(22), 2245-2256.*

Peul, W.C., van den Hout, W.B., Brand, R., Thomeer, R.T., Koes, B.W., Leiden-The Hague Spine Intervention Prognostic Study Group (2008). Prolonged conservative care versus early surgery in patients with sciatica caused by lumbar disc herniation: two year results of a randomised controlled trial *BMJ; 336(7657): 1355–1358.*

Pocock, S.J. (1983). *Clinical Trials: A Practical Approach.* Wiley

Schaafstra, A., Spinnewijn, W., Bons, S., Borg, M., Koes, B., Ostelo, R., Spijker-¬Huiges, A., Burgers, J., Bouma, M., Verburg,

A. (2015). *NHG-Standaard Lumbosacraal radiculair syndroom (Tweede herziening). Huisarts Wetenschap, 58(6),308-20.*

Soer, R., Reneman, M.F., Speijer, B.L., Coppes, M.H., Vroomen, P.C. (2012). Clinimetric properties of the EuroQol-5D in patients with chronic low back pain. *Spine Journal,12(11),1035-1039.*

Stafford, M.A., Peng, P., Hill, D.A. (2007). Sciatica: a review of history, epidemiology, pathogenesis, and the role of epidural steroid injection in management. *British Journal of Anaesthesia, 99(4), 461-473.*

Twisk, J.W.R. (2006). *Applied multilevel analysis: a practical guide for medical researchers.* Cambridge University Press.

van Helvoirt, H., Apeldoorn, A.T., Ostelo, R.W., Knol, D.L., Arts, M.P., Kamper, S.J., van Tulder, M.W. (2014).Transforaminal epidural steroid injections followed by Mechanical Diagnosis and Therapy to prevent surgery for lumbar disc surgery. *Pain Med.,15(7),1100-1108.*

Versteegh, M.M., Vermeulen, K.M., Evers, S.M., de Wit, G.A., Prenger, R., Stolk, E.A. (2016). Dutch Tariff for the Five-Level Version of EQ-5D. *Value in Health, 19, 343-352.*

Weinstein, J.N., Lurie, J.D., Olson, P.R., Bronner, K.K., Fisher, E.S. (2006). United States' trends and regional variations in lumbar spine surgery:1992-2003. *Spine,31(23),2707-2714.*

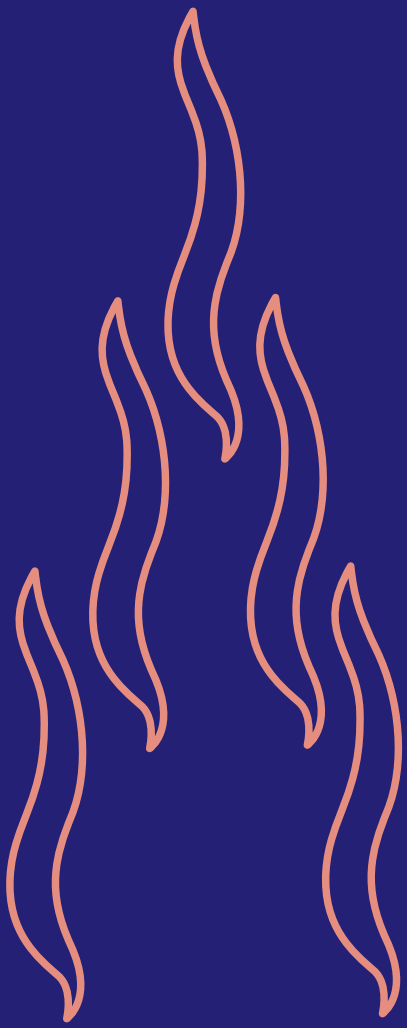
White, I.R., Royston, P., Wood, A.M. (2011). Multiple imputation using chained equations: Issues and guidance for practice. *Statist Med., 30, 377-399.*

Williamson, A., Hoggart, B. (2005). Pain: a review of three commonly used pain rating scales. *Journal of Clinical Nursing,14(7), 798–804.*

Xuemei, L., George, M.L., Kakouras, I., Edwards, C.L., Pietrobon, R., Richardson, W., Hey, L. (2003). Reliability, Validity, and Responsiveness of the Short Form 12-Item Survey (SF-12) in

Patients With Back Pain. *Spine*, *28(15),1739-1745.*

Yamato, T.P., Maher, C.G., Saragiotto, B.T., Catley, M.J., McAuley, J.H. (2017). The Roland–Morris Disability Questionnaire: one or more dimensions? *European Spine Journal*, *26(2)*, *301-308.*



5

Effectiveness and cost-effectiveness of combination therapy (Mechanical diagnosis & treatment & Transforaminal Epidural Steroid Injections) for patients with an indication for a lumbar herniated disc surgery: A randomized controlled trial

Mutubuki EN[1],*, van Helvoirt H[2], van Dongen JM[1], Ben Jornada A[3], Vleggeert-Lankamp CLA[4], Huygen F[5], van Tulder MW[1], Klopper-Kes AHJ[2], Rooker S[6], Leanders M[7], van der Gaag NA[4,7], Hoffmann CFE[8,9], Leilifeld P[10], Ostelo RWJG[1,12] , on behalf of the PLUS Study Research Group

[1] Department of Health Sciences, Faculty of Science, Vrije Universiteit Amsterdam, Amsterdam Movement Sciences, the Netherlands
[2] Stichting Rugpoli, Delden, the Netherlands
[3] Department of Health Sciences, Faculty of Science, Vrije Universiteit Amsterdam, the Netherlands
[4] Department of Neurosurgery, Leiden University Medical Centre, Leiden, the Netherlands
[5] Department of Anesthesiology, Centre of Pain Medicine, Erasmusmc, Rotterdam,
[6] Department of Neurosurgery, Kliniek ViaSana, Mill, The Netherlands
[7] Department of Neurosurgery Medische Spectrum Twente, Enschede, the Netherlands
[8] Department of Neurosurgery, HagaZiekenhuis, Den Haag, the Netherlands
[9] Department of Neurosurgery, Reinier de Graaf Gasthuis, Delft, the Netherlands
[10] Department of Neurosurgery, Amphia, Breda, the Netherlands
[11] Department of Neurosurgery, Canisius Wilhelmina Ziekenhuis, Nijmegen, the Netherlands
[12] Department of Epidemiology and Biostatistics, Amsterdam UMC, Amsterdam Movement Sciences, the Netherlands

This preliminary analysis will be incorporated in the paper describing the final analysis

LIST OF COLLABORATORS

Spaarne Gasthuis	S. Vissers	Svissers2@spaarnegasthuis.nl
HagaZiekenhuis	N. van der Gaag	N.vanderGaag@hagaziekenhuis.nl
	Rachel Tjoa-Bakker	r.tjoa@hagaziekenhuis.nl
	Jorrit de Vries	jo.devries@hagaziekenhuis.nl
Reinier de Graaf ziekenhuis	C.F.E. Hoffmann	c.hoffmann@hagaziekenhuis.nl
	C. Hegge	C.Hegge@rdgg.nl
	F. Kloet	f.kloet@haaglandenmc.nl
Medisch Spectrum Twente	M. Lenders	M.Lenders@mst.nl
	M. Abbink	m.abbink@mst.nl
	K. Kho	k.kho@mst.nl
	N. Hoess	n.hoess@mst.nl
Canisius Wilhelmina Ziekenhuis	Martine Van Bilsen	Martine.vanBilsen@radboudumc.nl
Kliniek ViaSana	Y. Pronk	Y.Pronk@viasana.nl
	S.Rooker	S.Rooker@viasana.nl
Amphia ziekenhuis	K. Kraus	Kkraus@Amphia.nl
	P. Leliefeld	PLeliefeld@amphia.nl
Rugpoli	Elmar Kleinjan	E.Kleinjan@rugpoli.nl
	Marije Pol	M.Pol@rugpoli.nl

ABSTRACT

Objectives: This study aimed to evaluate the effectiveness and cost-effectiveness of combination therapy, consisting of Mechanical Diagnosis and Treatment and Transforaminal Epidural Steroid Injections, compared to no intervention while being on the waiting list for lumbar herniated disc surgery. The study aimed to include 146 participants; this chapter reports a preliminary analysis of 56 patients with 6 months follow-up until the beginning November 2019. Both the effectiveness and cost-effectiveness analyses are presented.

Methods: Patients were randomly assigned to combination therapy (intervention group)(n=27) or to the waiting list for surgery (control group)(n=29). The primary outcome was the number of patients undergoing lumbar disc surgery during 6-month follow-up. Secondary outcomes included back and leg pain intensity (NPRS), physical functioning (RMDQ-23), self-perceived recovery (GPE), and health-related quality of life (EQ-5D-5L). For the economic evaluation, total societal and total healthcare costs during 6-month follow-up were measured. The combination therapy's effectiveness compared with usual care was assessed using mixed models (intention-to-treat) and its cost-effectiveness was assessed using seemingly unrelated regression analyses.

Results: A total of 24 out of 29 patients (83%) received surgery in the control group, whereas 9 out of 27 patients (33%) received surgery in the intervention group. The adjusted odds ratio of receiving surgery in the intervention group was 0.07 (95%CI: 0.02 to 0.35) compared to the control group. Converting this odds ratio to a relative risk (RR), results in a RR of 0.35 (95%CI: 0.13 to 0.79). There were no statistically significant differences in clinical effects between both groups. Costs were on average lower by €2,878 in the intervention group compared to the control group. For surgery, the ICER was 1,363, meaning that on average €1,363 were saved per surgery prevented in the intervention group compared to the control group.

Conclusion: This preliminary analysis indicates that combination therapy while being on the waiting list for lumbar herniated disc surgery might prevent such surgeries from happening compared

with no intervention. Apart from the difference in surgeries prevented, both groups had similar clinical effects, and the combination therapy's cost-effectiveness seems promising. However, we do not know what will happen to the results when the follow-up duration is extended to 12 months and when 146 patients are included, as planned.

Keywords: effectiveness, cost-effectiveness, lumbar disc herniation, lumbar disc surgery, mechanical diagnosis therapy, steroid injections

INTRODUCTION

Sciatica, also known as lumbosacral radicular syndrome, is commonly caused by a lumbar disc herniation. Sciatica it characterised by nerve pain that radiates from the buttock downwards along the course of the sciatica nerve[1] and is oftentimes accompanied by sensory and/or motor deficits[2]. The pathophysiology of sciatica is not only attributed to pressure on the nerve roots, but also to a complex interplay of inflammatory, immunological, and pressure-related processes[3]. Nonetheless, the primary source of pain is from the impingement of the nerve root, which causes about 85% of sciatica cases[4]. The lifetime prevalence of sciatica ranges from 12% to 43% and its point prevalence ranges from 1.6% to 13.4%[5-9]. The estimated incidence rate of sciatica in Western countries is 5 cases per 1000 adults[10]. In the Netherlands, the incidence rate of sciatica is slightly higher, being 9.4 cases per 1,000 patients per year[8]. Men are reported to be more affected than women, and so is the 45 to 64 years age group.

Differences in diagnosis and management of sciatica exist within and between countries[11]. Diagnosis of sciatica is based mainly on the patient's history and neurological examination[12], and its management is conservative in the acute stages. Conservative treatment includes provision of information regarding sciatica, advise to stay active, and pain management[12]. If complaints persist and no improvement is noted following conservative treatment, guidelines recommend referral to a spine surgeon who will evaluate if there is an indication for surgery[12, 13]. Despite similar recommendations in different national guidelines, there are international disparities in spinal surgery rates[14, 15]. To illustrate, spinal surgery rates in the United States are 30% higher than in the Netherlands, 80% higher than in the UK, and 50-60% higher than in Canada[10]. An estimated 8,000 sciatica surgeries were performed in 2018 in the Netherlands[16].

The natural course of recovery of sciatica is in most cases favourable[17, 18], however no distinction can be made between patients who will experience a favorable natural course of recovery and those that could benefit from early surgical

consultation[19]. Surgical recovery rates vary and some patients still suffer residual complaints after two years[20]. Repeated surgical rates are between 10-15% within two years[18]. Surgery may reduce pain significantly in the short and medium-term compared to conservative therapy[12, 21, 22]. However, there are no differences in health-related quality of life and small to negligible differences in physical function and disability compared with non-surgical interventions[21]. In the long-term, surgical and conservative treatment of lumbar hernia are equally successful[12, 23]. Epidural steroid injections are a commonly used conservative therapy for sciatica[24] [25]. At short-term follow-up, epidural steroid injections slightly reduce leg pain and disability in persons with sciatica[26, 27],[25]. Mechanical diagnosis and therapy (MDT) is also used as conservative treatment in sciatica patients[28]. MDT is a way of assessing and treating spinal and extremity pain. It is based on responses to repeated end range movements, which determine the direction of preference[29]. At short-term, MDT compared to advice only to stay active, was found to reduce low back pain significantly and at medium-term, leg pain and low back pain reduced significantly more[30].

Both MDT and epidural steroid injections have the potential to reduce pain and disability in sciatica patients and hence prevent surgery. However, their combined influence in reducing the amount of sciatica surgeries has only been evaluated in a pilot study by van Helvoirt et al[28]. The pilot study found that combining MDT and transforaminal epidural steroid injections (TESIs) has the potential to reduce the number of lumbar herniated disc surgeries, as only 22% of patients with a herniated lumbar disc and an indication for surgery, needed surgery after one-year follow-up[28]. However, the pilot study, including having less carefully selected surgical patients, lacked a control group, and hence patients could not be randomized across study conditions, and the cost-effectiveness of the intervention was not assessed. Therefore, this multicenter trial (i.e. the PLUS study) aimed to assess if a combination therapy (MDT and TESIs), administered while being on the waiting list for a lumbar herniated disc surgery, is effective and cost-effective compared to no intervention (while being on the waiting list) among patients with an MRI-confirmed indication for a lumbar

herniated disc surgery. In this thesis, we conducted a preliminary analysis, because patient recruitment lacked behind and we wanted to have a first indication about the effectiveness and cost-effectiveness of combination therapy versus usual care.

METHODS

Design

In this preliminary analysis, both an effectiveness analysis and an economic evaluation were performed using all 6-month follow-up data of the PLUS-study available at the beginning of November 2019. A detailed description of the PLUS-study, including its sample size calculation, can be found elsewhere[31].

In brief, the PLUS-study is a 12-month randomized controlled trial comparing a combination therapy, performed while being on the waiting list for lumbar herniated disc surgery, to no intervention while being on the waiting list (i.e. usual care) among sciatica patients. Since December 2017, patients were recruited from seven hospitals in the Netherlands; i.e. Kliniek ViaSana, Spaarne Gasthuis, Medische Spectrum Twente, Amphia ziekenhuis, Reinier de Graaf, Hagaziekenhuis and Canisius Wilhemina Ziekenhuis. Written informed consent was obtained from all patients. Following confirmation of eligibility and baseline assessments, patients were randomised to combination therapy or usual care by an independent researcher using a randomisation list that was created by a web-based randomisation program. Randomisation was performed at the individual level, and stratified on duration of complaints (i.e. <6 months vs. ≥6 months) and hospital. Blinding of the patients and treatment providers was practically impossible due to the nature of the intervention. The outcomes assessor was not blinded, because all outcomes were self-reported. Data analyses were performed by a researcher who was not involved in the treatment and measurement of patients. The Medical Ethics Committee of the VU University Medical Centre in Amsterdam approved the study protocol (NL60558.029.17) and all study procedures were approved by the Board of Directors of participating

centers. The study protocol was registered in the Dutch Trial Register (NTR6715) and in the European Union Drug Regulating Authorities Clinical Database (EudraCT number: 2017-002119-33).

Participants

Patients were eligible if they were aged 18 and above, had an indication for lumbar disc hernia surgery, had incapacitating sciatica with leg pain (NRS>6; with or without back pain) that had lasted for a minimum of 6 weeks with or without mild neurological deficit (i.e. Medical Research Council [MRC]>3), and had an MRI-confirmed herniated nucleus pulpous that compromised the spinal nerve and could explain their clinical symptoms[32]. Exclusion criteria were: spinal surgery and/or TESIs at the same level during the previous 6 months; bony stenosis; cauda equina syndrome; spondylolisthesis; pregnancy; complicated disc herniation requiring more than one operation; severe coexisting disease (e.g., osteoporosis and dementia); contraindications for TESIs; insufficient knowledge of the Dutch language; emergency surgery as determined by the neurosurgeon; and/or being allergic for Iohexol 240 mg/ml (i.e., OMNIPAQUE 240).

Treatments

Intervention condition: Combination therapy
The combination therapy had two parts, i.e. MDT and TESIs, and was delivered by teams of pain interventionists and physiotherapists at four outpatient clinics (in Dutch 'Rugpoli') throughout the Netherlands. Patients were first assessed by a pain interventionist for contraindications for TESIs. Then, they were classified as "centralizers" or "non-centralizers", using MDT principles, that is, assessment of the patients' pain pattern responses on repeated movement tests. Centralizers are patients with a centralization pain pattern, meaning that they experienced a clear change in leg pain location from a more peripheral location towards a more central location[33]. Centralizers received advice and treatment according to these principles with direction-specific exercises and posture

correction. Non-centralizers are patients with peripheralization pain pattern, meaning that they experienced a clear change in leg pain location towards a more peripheral location. Non-centralizers received TESIs. If pain reduction was less than 80%, a second TESI was administered with patient consent. A maximum of three TESIs were given to optimize pain relief. After the injections, the non-centralizers were re-classified using the aforementioned principles, into four sub-groups;

1) Resolved symptoms (i.e. no or irrelevant pain; ≤1 on a 0-10 NRS),
2) Centralizing and significantly less pain (i.e. a pain reduction of ≥2 on a 0-10 NRS),
3) Non-centralizing and significantly less pain (i.e. a pain reduction of ≥2 on a 0-10 NRS),
4) Non-centralizing with high levels of pain and disability (i.e. a pain score of ≥8 on a 0-10 NRS and a disability score of >10 on the RDMQ-23).

Patients in the first three sub-groups, received specific MDT exercises in one to six sessions, during a period of on average four weeks. Patients requiring more sessions were referred to accredited MDT therapists within the Rugpoli network. Patients in subgroup 4) were referred back to the neurosurgeon who assessed whether patients still required surgery.

Control condition: waiting list for surgery (usual care)
Control group patients received no intervention while being on a waiting list scheduled to receive lumbar disc surgery.

Use of co-intervention
Use of co-interventions by patients in both study conditions was allowed and monitored.

Effect measures

Primary outcome measure
The primary outcome measure was the proportion of patients undergoing lumbar surgery and in the current analyses we focus on the 6-month follow-up. Patients were scored as either having had a lumbar surgery or not. For this purpose, patients were asked whether they had a lumbar surgery since the previous

measurement point using an online questionnaire at 2, 4, and 6 months after randomisation, and the provided information was validated using hospital records.

Secondary outcome measures

In line with the core outcome set for clinical research and clinical practice (COS)[34], secondary study parameters included back and leg pain, physical functioning, and health-related quality of life. Additionally, self-perceived recovery was measured. Measurements took place at baseline, 2, 4, and 6 months after randomization using online questionnaires. An additional leg and back pain intensity measurement was carried out at 4 weeks after randomisation, the rationale is that, especially leg pain will improve in the combination group during the first weeks after treatment. Back and leg pain were measured using the Numeric Pain Rating Scale (NPRS). The NPRS ranges from 0 ("no pain") to 10 ("representing worst pain imaginable"). The numerical pain rating scale is reliable, valid and has good sensitivity[35].

Functional status was measured using the Roland Morris Disability Questionnaires (RMDQ-23)[36]. The RMDQ-23 consists of 23 "yes" or "no" questions, measuring limitation in activity associated with back and leg pain[37]. The scores range from 0 (no disability) to 23 (maximal disability)[37]. The RMDQ-23 is regarded as reliable and valid[38].

Self-perceived recovery was measured using the global perceived effect scale (GPE), a 7-point Likert scale ranging "from worse than ever" (1) to "completely recovered" (7). The responses were dichotomized into successful recovery (1-2) and non–successful recovery (3-7). Health-related quality of life was measured using the EQ-5D-5L. The EQ-5D-5L consists of five health dimensions; mobility, self-care, anxiety/depression, pain/discomfort, and daily activities. Patients are asked to rate their level of severity as; no, slight, moderate, severe problems, unable to perform[39]. The patients' EQ-5D-5L health states were converted into utility scores, ranging from 0 ("death") to 1 ("optimal health"), using the Dutch 5L value set[40]. Quality adjusted life years (QALYs) were calculated by multiplying the patients' time spent in a certain state by the respective utility value.

Cost outcome measures

Costs were assessed from a societal as well as a healthcare perspective. From the societal perspective, intervention, other healthcare, unpaid productivity, informal care, absenteeism, and presenteeism costs were included, whereas only costs accruing to the formal Dutch health sector were included when the healthcare perspective was applied. Resource use data were collected at baseline, 2, 4, and 6 months after randomisation using online cost questionnaires. Due to the 6 months follow-up, discounting of costs was not necessary[41].

All costs were converted to 2018 Euros using consumer price indices[42].

Intervention costs were micro-costed. That is, intervention costs were estimated by determining the number of attended MDT sessions and/or the number of received TESIs, and valuing them using their respected unit prices derived from 'Rugpoli' accounting records.

Other healthcare costs included use of primary or secondary healthcare and the use of medication; prescribed and over the counter. The use of primary and secondary healthcare was valued using the Dutch Manual of costing[43] and the use of medication was valued using prices derived from the website of the Dutch Health Care Institute (http://www.medicijnkosten.nl). For less common healthcare services, hospital accounting records and/or prices of professional organizations were used.

Informal care by family or other people (i.e. volunteers, friends) was valued using a recommended Dutch shadow price of €14,57 per hour (in Euros 2018)[43].

Unpaid productivity losses (i.e. losses in studies, voluntary work, domestic work) were valued using the same recommended Dutch shadow price of €14,57 per hour[43].

Absenteeism from paid employment was measured using the Productivity and Disease Questionnaire (PRODISQ)[44], and was valued in accordance with the Friction Cost Approach (FCA). Patients were asked to report the total number of days they were absent from work due to sickness. In accordance to the FCA, which assumes that productivity costs are limited to the friction period (i.e. period required to replace sick employee), absenteeism costs were truncated at a friction period of 12 weeks if necessary. Absenteeism costs were valued using gender-

specific price weights[45].
Presenteeism costs were estimated using an item of ''The World Health Organization Health and Work Performance Questionnaire'' (WHO-HPQ). This item consists of an 11-point scale, ranging from ''worst performance'' (0) to ''best performance'' (10). Participants were asked to rate their overall work performance during the previous two months (Wown). In the WHO-HPQ, presenteeism is defined as a measure of actual work performance in relation to ''best performance'' (10), irrespective of the presence or absence of health complaints. Hence, the participants' level of presenteeism per follow-up period was calculated using the formula below:
Level of presenteeism = (10 – Wown)/10
Subsequently, presenteeism days were calculated by multiplying the participants number of days worked (i.e. working days – absenteeism days) by their level of presenteeism. Presenteeism days were valued using valued using gender-specific price weights[45].

Data analysis

Analyses were performed according to the intention to treat principle. Baseline characteristics were compared between the two treatment groups using descriptive statistics. Missing data were imputed using Multivariate Imputation by Chained Equations with Predictive Mean Matching[46], and this procedure was stratified by treatment group. Missing cost and effect data were assumed to be Missing At Random, meaning that missing data depends only on observed data and not on unobserved data[47] and that a systematic difference might exist between the missing and observed data, however these can be explained entirely by other observed variables[48]. Included in the imputation model were variables related to the "missingness" of data or associated to the outcome and variables that differed at baseline. These variables included number of sick days, medication, and the patients' treatment expectations for combination therapy and surgery. A total of 10 data sets were created to reach a loss of efficiency lower than 5%. Pooled estimates were calculated according to Rubin's rules, thereby including both within and between imputation variability[49].

The primary outcome (i.e. surgery; yes/no) was analysed

using a logistic regression analysis, which was corrected for possible confounders. Baseline variables changing the regression coefficient by more than 10% were deemed confounders and were included in the model. Odds ratios (ORs) with 95% confidence intervals (95%CIs) were calculated and converted to relative risks (RRs) using the method of Zhang et al (1998): RR = OR/[(1 − prevalence in control group) + (prevalence in control group × OR)][50]. Secondary outcomes were analysed using linear (for continuous outcomes) and logistic (for dichotomous outcomes) mixed models for repeated measurements at 4 weeks (only for pain), and 2, 4, and 6 months, corrected for possible confounders. For analysing the secondary outcomes, data were not imputed, as previous research indicated that in longitudinal data analysis, multiple imputation of missing values is not necessary, regardless of the missing data mechanism[51].

For the economic evaluation, seemingly unrelated regression (SUR) analyses were performed to estimate total cost and effect differences. With SUR analyses, two separate regression models are estimated simultaneously (i.e. one for costs and one for QALYs) and the correlation between costs and QALYs is accounted for through correlated error terms. Cost and effect differences were adjusted for confounders. Additionally, the difference in QALYs was adjusted for the patients' baseline utility score[52].

Incremental cost-effectiveness ratios (ICERs) were calculated by dividing the difference in costs by that in effects. Bias-corrected and accelerated bootstrapping with 5000 replications were used to estimate the uncertainty surrounding the ICERs and 95%CIs around cost differences. Uncertainty was shown by plotting bootstrapped incremental cost-effect pairs (CE pairs) on cost-effectiveness planes (CE-planes)[53-55]. Cost-effectiveness acceptability curves (CEACs) were created to show the joint uncertainty surrounding costs and effects. CEACs indicate the probability of combination therapy (intervention) being cost-effective compared with no intervention (control) for a range of ceiling ratios (i.e. the maximum amount of money decision-makers are willing to pay per unit of effect gained).
All analyses were performed in STATA (V16). A p-value of <0.05 was considered statistically significant.

Sensitivity analysis

Two sensitivity analyses were performed. The first sensitivity analysis used only data of patients with complete cost and effect values. In the second sensitivity analysis, the Human Capital Approach (HCA) was used to estimate absenteeism costs. In the HCA approach, all productivity hours lost are included in the absenteeism cost estimate, instead of only those occurring during the friction period.

RESULTS

Population

Data from 56 patients, recruited from 7 hospitals in the Netherlands were used for this preliminary analysis. The baseline characteristics of the two study groups are presented in Table 1. There were more males (62%) in the control group than in the intervention group (52%). In addition, 17% of people had a low education level in the control group compared to 7% in the intervention group. No comorbidities were reported in the control group, whereas 3 persons (11%) reported having comorbidities in the intervention group. The mean duration of complaints was higher in the control group compared with the intervention group, i.e. 15 months versus 6 months. The flow of patients in the study is shown in the flow diagram (Figure 1).

Effectiveness

Primary outcome

A total of 24 out of 29 patients (83%) received surgery in the control group, whereas 9 out of 27 patients (33%) in the intervention group received surgery, resulting in an unadjusted risk difference of 50%. The corresponding adjusted odds ratio of receiving surgery in the intervention group was 0.07 (95%CI: 0.02 to 0.35) compared to the control group (Table 2). That is, the odds of surgery was much lower in the intervention group compared to control group. The odds ratio was converted to relative risk (RR), resulting in a RR of 0.35 (95%CI: 0.13 to 0.79).

Secondary outcomes

Table 2 shows that there were no statistically significant differences between both groups for any of the secondary

outcomes. Back pain intensity reduced in both groups; i.e. from 5.89 at baseline to 3.65 at 6 months in the intervention group and from 6.00 at baseline to 2.23 at 6 months in the control group. The corresponding difference in back pain between both groups during follow-up was small, in favour of the control group, and not statistically significant (MD 0.21; 95%CI: 0.01 to 0.41). Leg pain intensity reduced in both groups; i.e. from 7.67 at baseline to 2.50 at 6 months in the intervention group and from 6.76 at baseline to 2.38 at 6 months in the control group. The corresponding difference in leg pain between both groups during follow-up was small, in favour of the control group, and not statistically significant (MD -0.03; 95%CI:-0.26 to 0.21). Physical functioning improved in both groups; i.e. from 16.2 at baseline to 7.9 at 6 months in the intervention group and from 16.5 at baseline to 5.8 at 6 months in the control group. The corresponding difference in physical function between both groups during follow-up was small, in favour of the control group, and not statistically significant (MD 0.33; 95%CI:-0.144 to 0.82). For GPE, there were no statistically significant differences between both groups (Table 2).

Costs

Table 3 shows the cost differences between study groups during 6 months follow-up and mean costs per participant in the intervention and control groups. The mean total intervention costs were €692 per patient. Costs of presenteeism, total societal costs, and total healthcare costs were lower in the intervention group compared with the control group, but not statistically significant. The mean total surgery costs were statistically significantly higher in the control group (€5,182 per patient) compared to the intervention group (€2,296 per patient), (cost difference: -3,048; 95%CI: -4,626 to -1,259). Medication costs, on the other hand, were statistically significantly higher in the intervention group (€135) compared to the control group (€74), (cost difference: 61; 95%CI: 16 to 132). Healthcare costs were the highest contributor of total societal costs in both the intervention and control groups.

Cost-effectiveness: societal perspective

The analysis from the societal perspective (Table 4) showed that the ICER for QALYs was 820,404. This ICER indicates that the

costs were €820,404 lower in the intervention group per QALY lost compared with the control group. Thus, on average, the intervention was less costly, but also less effective. Please be aware that the large ICER was due to the small effect on QALYs (MD: -0.004; 95%CI: -0.051 to 0.043). The CEAC in Figure 3a shows that at a willingness to pay of 0, 20,000, and 80,000 Euros per QALY gained, the probability of the intervention being cost-effective compared to usual care is 0.8, 0.8, and 0.7, respectively.

For surgery, the ICER was 1,363, meaning that on average €1,363 were saved per surgery prevented in the intervention group compared to the control group. Hence, combination therapy dominates usual care. Please note that surgery costs were not included in the total societal cost estimate of this analysis to prevent double counting. The CEAC in Figure 3b, shows that at a willingness to pay of 0, 20,000 and 80,000 euros per surgery prevented, the probability of the intervention being cost-effectiveness compared to usual care is 0.6, 0.9, and 1, respectively.

Cost-effectiveness: healthcare perspective

The analysis from the healthcare perspective (Table 4) showed that the ICER for QALYs was 2,918,244. This ICER indicates that the costs were €2,918,244 lower in the intervention group per QALY lost compared with the control group. Thus on average, the intervention was less costly, but also less effective. Again, the large ICER was due to the small effect on QALYs (MD: -0.001; 95%CI: -0.047 to 0.045). The CEAC showed that at a willingness to pay of 0, 20 000, and 80 000 euros per QALY gained, the probability of the intervention being cost-effectiveness compared to usual care is 0.8, 0.8, and 0.8 respectively.

For surgery, the ICER was -1,713, meaning that on average €-1,713 was spent per surgery prevented in the intervention group compared to the control group. The CEAC, shows that at a willingness to pay of 0, 20 000, and 80 000 euros per surgery prevented, the probability of the intervention being cost-effectiveness compared to usual care is 0.2, 1 and 1, respectively.

Sensitivity analysis

When analysing only the complete cases (SA1), in comparison to

the main analysis, the difference in total costs was smaller from the societal perspective (i.e. -1575 versus -3164 for QALYs) and even became in favour of the control group when surgery costs were excluded (1727 versus -645 for surgery). When using the HCA instead of the FCA (SA2), the difference in total societal costs was slightly smaller (e.g. -2883 versus – 3164 for QALYs). In line with the main analysis, differences in QALYs were in favour of the control group and not statistically significant in both sensitivity analyses. Overall, the conclusions of the study would not change when using the results of any of the sensitivity analyses.

DISCUSSION

Main findings

This study investigated the effectiveness and cost-effectiveness of a combination therapy while being on the waiting list for a lumbar herniated disc surgery versus being on the waiting list for lumbar disc herniation surgery alone. Surgeries were found to be prevented in the intervention group compared with the control group and there were no differences in clinical effects between both groups. The results showed that the odds ratio of receiving surgery in the intervention group compared with the control group was 0.07 and statistically significant. The odds ratio appears to be big, and this is because the prevalence of the outcome in the control group is high. Therefore, the odds ratio was also converted to a relative risk (RR), resulting in a RR of 0.35 (95%CI: 0.13 to 0.79). This RR indicates that the "risk" of receiving surgery was about one-third in the intervention group (i.e. combination therapy) of that in the control group (i.e. waiting list for lumbar disc surgery).

The results also showed substantial differences between both groups for presenteeism, total societal, and healthcare costs, but these differences were not statistically significant. The latter, however, is likely due to the highly skewed nature of cost data, resulting in them requiring very large sample sizes to detect relevant between group differences. Moreover, if decision-makers are not willing to pay anything for preventing

a surgery, the probability of combination therapy being cost-effective compared with usual care was 0.6 and this probability increased with increasing values of willingness to pay to 1.0 at a willingness to pay of €20,000. It is currently unknown what decision-makers are willing to pay per prevented surgery. Thus, strong conclusions about the combination therapy's cost-effectiveness for this outcome cannot be made. For QALY, willingness to pay thresholds are available for various countries. At the lower bound of the Dutch willingness to pay threshold for QALYs, i.e. €10,000/QALY, combination therapy had a probability of being cost-effective compared to control of about 0.81. At the NICE willingness to pay threshold of £20,000 (about €23,000) the probability of the intervention being cost-effective was 0.79 per QALY gained. Results of the sensitivity analyses were more or less in line with the results of the main analysis. All in all, this preliminary analysis indicates that combination therapy while being on the waiting list for lumbar herniated disc surgery might prevent such surgeries compared with usual care, with no differences between groups in terms of clinical outcomes and its cost-effectiveness seems promising.

In both the intervention and control group, the percentage of patients that reported to be successfully recovered on the GPE scale at 6 months follow-up was low, i.e. 53% in the intervention group and 42% in the control group. In the study by Peul et al. (2007), recovery rates were much higher, i.e. 95% for both groups. This might be explained by the longer follow-up of Peul et al. (2007)(i.e. 1-2 years), during which natural recovery has likely taken place. The follow-up in our study was only 6 months, hence natural recovery might have not fully taken place. Therefore, although the combination therapy's cost-effectiveness seems promising, this is a preliminary analysis and firm conclusions should not be drawn before data of more participants and at a longer term follow-up are available. In addition, the GPE scale does not seem to accurately measure change over time, as it has previously been found to be strongly influenced by current health status[56]. Therefore, it might also be that the PLUS-study patients did not take into account their baseline health status, but their current health status, when scoring GPE.

Comparison with literature

The results of the current study show that 83% of the patients received surgery in the control group compared to 33% in the intervention group, highlighting that the combination therapy is indeed promising in reducing the amount of surgeries. The pilot study by Helvoirt et al. (2014), which preceded the current study, found that only 22% of sciatica patients required surgery after receiving the combination therapy. Hence, the pilot study reported lower amounts of hernia surgeries performed (22%) compared to the current study (33%). However, the selection of patients in the PLUS-study might have been more careful than in the pilot study, as less patients are currently referred for surgery compared to a decade ago when data were collected for the pilot study. Hence, the patients referred for the PLUS-study were probably more likely to truly have an indication for surgery, thereby decreasing the likelihood that conservative therapy might be successful.

Epidural steroid injections provide modest pain relief in the short-term in patients with lumbar disc hernia[57, 58] and were reported to have no impact on the incidence of surgery or physical disability[57]. In contrast, the present study showed that the incidence rate of surgeries could be prevented when TESIs are combined with MDT. However, no significant differences regarding physical function, pain, and perceived recovery (secondary outcomes) between the intervention and control group were observed. This is in line with the results of Pennington et al. who also reported no statistical significant difference in patient-related outcomes when epidural steroid injections were compared to medical management alone[58].The combination therapy's main effect is prevention of surgery, which does not seem to be mediated by an effect of the intervention on pain and disability. The short-term effects (i.e. 4 weeks and 2 months) on pain and physical functioning were similar in both groups, suggesting that participants on the waiting list who didn't receive any treatment improved as much as participants who received the combination therapy group during the first weeks. The effect on the primary outcome in the combination therapy group might be triggered by other intervention-related factors, such as the provision of information regarding low back pain, self-assurance, and validation of patient complaints by

the therapists/physician or the effects may reflect the natural course natural. This reasoning is underscored by the fact that there was a clear difference regarding the support (from the physiotherapist) that patient received in both groups, which may have caused more patients to abstain from surgery in the intervention group compared with the control group.

To date, with exception to the current study, no full economic evaluation has compared the combination therapy (MDT &TESIs) to usual care for patients with sciatica on a waiting list for lumbar disc surgery. A recent systematic review reported that the cost-effectiveness of surgery compared to conservative treatment depends on the willingness to pay threshold[21]. Although the strength of the evidence was very low, the mean cost per QALY gained from the payer perspective ranged from $51,156 to $83,322 united states dollars in 2010[21]. These values are, however, higher than the lower bound of the Dutch willingness to pay threshold (i.e. €10 000/QALY) and the NICE threshold (i.e. £20 000 about €23 000).

Strengths and limitations

The use of a societal as well as a healthcare perspective was an important strength of this study. An important advantage of using the societal perspective is that it assesses the benefits from investments in all sectors of the economy[59]. The healthcare perspective, on the other hand, may assist healthcare decision-makers more in achieving healthcare goals than the societal perspective[60]. Another strength of the present study is its pragmatic RCT design, which is regarded as ideal for effectiveness analyses and economic evaluations, because it allows evaluation of an intervention's effectiveness and cost-effectiveness under real-life conditions, improving external validity, while enabling collection of cost and effect data prospectively and randomization improving internal validity. However, external validity might have been decreased in this study due to a slow recruitment rate. Hence, it is uncertain whether the included participants are a good representation of all patients waiting for surgery. Another strength is the use of multiple imputation to deal with missing data in the primary effect analysis and the economic evaluation, thus avoidance of a complete-case analysis, which reduces a study's power and precision. Although multiple

imputation is the preferred method of dealing with missing data in economic evaluations, it relies on the assumption that data are missing at random, meaning that missing data depends only on observed data and not on unobserved data. This assumption is not always true; hence, care should be exercised in interpretation of the results. Multiple imputation is not required when using mixed models and hence mixed models with maximum likelihood estimates (MLE) were used for analyzing the secondary outcomes without imputing missing values. Moreover, relevant statistical challenges in economic evaluations were accounted for using appropriate statistical analysis, i.e. SUR to account for possible correlation between effects and costs[61, 62], multiple imputation was used to deal with missing data[63-65], skewed data was dealt with using non-parametric bootstrapping[66-71] and baseline imbalances were accounted for using regression-based adjustment[52, 72, 73]. A failure to adequately account for statistical challenges may bias the results and consequently lead to invalid decisions and potential wastage of resources.

A limitation of this study is its relatively small sample size (n=56), which could have made it difficult to detect relevant effect differences between study groups. According to our sample size calculation, we needed to include 146 patients to detect a difference of 30% (i.e. a reduction of 30% of surgeries in the intervention group compared to usual care). As described above, the current results are based on a preliminary analysis. We are continuing recruitment and will report the final analysis in a peer-reviewed scientific journal. Another limitation of the present study is the relatively short follow-up period of 6 months, which might not be long enough given that LBP is an episodic condition, with a natural recovery component. Thus, it might be possible that patients that recovered at 6 months experience another similar complaints after the 6 months. Therefore, patients who did not receive surgery because they received the combination therapy (intervention) might still receive surgery due to recurrent LBP complaints after the 6 months follow-up in the current study, thereby decreasing the difference between groups in our main outcome. The final analysis will include 12 months follow-up data. Retrospective questionnaires were used to collect cost and effect data in this study and these are prone to recall bias and or social desirability bias. However, an attempt

to limit recall bias was done by shortening the recall period to about 2 months. To try and reduce social desirability bias, self-administration of the questionnaires were done in this study[74].

Conclusion

This preliminary analysis indicates that combination therapy while being on the waiting list for lumbar herniated disc surgery might prevent such surgeries from happening compared with usual care, that both have similar clinical effects, and that the combination therapy's cost-effectiveness seems promising. However, data of more participants and a longer follow-up duration are needed to establish this.

Table 1: Patient baseline characteristics

Participant characteristic	**Surgery** (control) (n=29)	**Combination therapy** (intervention) (n=27)	**PLUS-study** (n=56)
Age (years) [mean (SD)]	47 (11)	46 (15)	46 (13)
Gender [n (%)]			
Female	11 (38)	13 (48)	24 (43)
Male	18 (62)	14 (52)	32 (57)
Depressive feelings [n (%)]			
Yes	16 (55)	13 (48)	29 (52)
No	13 (45)	14 (52)	27 (48)
Anxiety feelings [n (%)]			
Yes	15 (52)	14 (52)	29 (52)
No	14 (48)	13 (48)	27 (48)
Educational level [n (%)]			
Low (no education, primary level education, lower vocational and lower secondary education)	*5 (17)*	*7 (26)*	*12 (21)*
Moderate (higher secondary education or undergraduate)	*18 (62)*	*12 (44)*	*30 (54)*
High (tertiary, university level, postgraduate)	*6 (20)*	*8 (30)*	*14 (25)*
Living together with a partner [n (%)]			
Yes	*22(76)*	*15 (56)*	*37 (66)*
No	*7(24)*	*12 (44)*	*19 (34)*
Nationality [n (%)]			
Dutch	26 (90)	25 (93)	51 (91)
Non-Dutch	3 (10)	2(7)	5(9)
Painkillers against leg/back pain [n (%)]			
Yes	23 (79)	21 (78)	44 (79)
No	6 (21)	6 (22)	12 (21)
Employment [n (%)]			
Yes	21 (72)	17 (68)	38 (68)
No	8 (28)	10 (32)	18 (32)
Comorbidities [n (%)			
Yes	0 (0)	3 (11)	3 (5)
No	53 (95)	24 (89)	53 (95)

Continued - Table 1: Patient baseline characteristics

Smoking [n (%)]			
1= Yes	14 (48)	14 (52)	28(50)
2= No	15 (52)	13 (48)	28(50)
Patients expectations [mean (SD)] range 0-100			
Combination therapy	52 (28)	63 (27)	57 (28)
Surgery	76 (21)	68 (21)	72 (22)
Duration complaints [mean (SD)] in months	15 (40)	6(2)	10 (29)
RMDQ 23 [mean (SD)]	31 (4)	32 (4)	32 (4)
Pain catastrophizing [n (%)]			
Yes	16 (55)	14 (52)	30 (54)
No	13 (45)	13 (48)	26 (46)
Utility [mean (SD)] range 0-1	0.4 (0.3)	0.4 (0.3)	0.4 (0.3)
Pain intensity [mean (SD)] range 0-10			
Leg	7 (2)	8 (2)	7 (2)
Back	6 (2)	6 (3)	6 (3)
Note: percentages have been rounded off			

Table 2: Treatment effects for primary and secondary outcomes

[a]Primary Outcome:				
	Odds ratio	**Lower bound**	**Upper bound**	**P-value**
Intervention group (ref: usual care)	0.07	0.02	0.35	0.001
[b]Secondary Outcomes:				
		Intervention group Mean (SD)	**Control group Mean (SD)**	**Treatment effect Mean Difference (95%CI)**
Outcome				
Back pain range 0-10	**Overall effect**			**0.21 (0.01 to 0.41)**
	Baseline	5.89 (2.76)	6.00 (2.45)	
	4 weeks	4.92 (2.30)	4.54 (2.90)	0.53 (-0.96 to 2.02)
	2 months	4.00 (0.48)	3.25 (0.51)	0.92 (-0.58 to 2.43)
	4 months	3.10 (2.53)	2.74 (20.7)	0.47 (-1.05 to 1.99)
	6 months	3.65 (2.70)	2.23 (2.12)	1.61 (0.07 to 3.16)
Leg pain range 0-10	**Overall effect**			**-0.03 (-0.26 to 0.21)**
	Baseline	7.67 (1.84)	6.76 (2.46)	
	4 weeks	5.38 (2.79)	4.77 (3.31)	-0.23 (-1.87 to 1.40)
	2 months	3.27 (3.24)	2.68 (2.48)	-0.16 (-1.81 to 1.49)
	4 months	2.57 (2.78)	2.07 (2.43)	-0.35 (-2.01 to 1.32)
	6 months	2.50 (2.50)	2.38 (2.53)	-0.69 (-2.38 to 0.99)
Physical functioning range 0-23	**Overall effect**			**0.33 (-0.144 to 0.82)**
	Baseline	16.2 (5.1)	16.5 (3.7)	-0.30 (-2.38 to 1.79)
	2 months	10.2 (7.2)	9.6 (6.8)	0.61 (-3.35 to 4.57)
	4 months	8.3 (6.7)	6.8 (6.1)	1.47 (-2.29 to 5.24)
	6 months	7.9 (6.7)	5.8 (6.9)	2.12 (-1.92 to 6.16)
		Intervention group N (%)	**Controlgroup (%) N (%)**	**Treatment effect Odds Ratio (95%CI)**
Global Perceived Effect Success 1&2 is success	**Overall effect**			**0.92 (0.60 to 1.42)**
	2 months	9 (41)	13 (46)	0.55 (0.37 to 8.17)
	4 months	10 (48)	15 (56)	0.77 (0.06 to 15.7)
	6 months	10 (53)	15 (42)	1.04 (0.07 to 15.7)

[a]Adjusted for employment
[b]Adjusted for age, sex, depressive feelings, anxiety feelings, education level, living together, nationality, medication (painkillers), smoking, pain catastrophizing, employment, patients expectations, comorbidities, duration complaints

Table 3: Cost differences between the groups during 6 months follow-up and mean cost per participant in the intervention and usual care

Cost category	Intervention n=27, mean (SEM)	Control n=29, mean (SEM)	Cost difference crude, mean (95%CI)	[c]Cost difference adjusted, mean (95%CI)
Intervention costs	692 (92)	0	692 (500 to 866)	698 (460 to 905)
Primary medical costs	2381 (616)	2613 (763)	-232 (-2257 to 1398)	25 (-2470 to 1789)
Secondary medical costs	453 (181)	605 (211)	-153 (-765 to 281)	-2 (-598 to 520)
Surgery costs	2296 (898)	5182 (615)	-3048 (-4626 to -1259)	-2878 (-4924 to -532)
Medication costs	135 (27)	74 (12)	61 (16 to 132)	69 (15 to 167)
Informal care costs	1368 (425)	1032 (338)	336 (-700 to 1408)	853 (-430 to 2285)
Absenteeism costs	3441 (1090)	5274 (1138)	-845 (-3327 to 1703)	-395 (-3009 to 2466)
Presenteeism costs	3311 (703)	5446 (1049)	-2135 (-4581 to 248)	-2006 (-4784 to 271)
Unpaid productivity costs	2381 (616)	2613 (763)	-232 (-2257 to 1398)	25 (-2470 to 1788)
Total societal cost	16458 (2361)	22839 (2631)	-6381 (-13126 to -86)	-3610 (-10 764 to 2867)
Total societal cost – without the cost of surgery	14161 (2191)	17657 (2687)	-3496 (-10278 to 2979)	-732 (-7943 to 5125)
Healthcare costs	5957 (1193)	8474 (907)	-2517 (-4994 to -20)	-2087 (-5077 to 1166)

[c]Adjusted for education level, living together, duration of complaints, amount of sick days, work presentation

Table 4: Differences in pooled mean costs and effects (95% Confidence intervals), incremental cost-effectiveness ratios, and the distribution of incremental cost-effect pairs around the quadrants of the cost-effectiveness planes

Analysis	**Sample Size**		**Outcome**	**ΔC (95% CI)**	**ΔE (95% CI)**	**ICER**	**Distribution CE-plane (%)**			
	Intervention	Control		€	Points	€/point	NE[1]	SE[2]	SW[3]	NW[4]
Main analysis –	27	29	QALYs (Range: 0 - 1)	-3164 (-9916 to 3213)	-0.004 (-0.051 to 0.043)	820 404	1.8	42.8	40.9	14.5
Imputed dataset	27	29	Surgery (Range: 0-1)	-645 (-7764 to 5314)	-0.473 (-0.736 to -0.209)	1363	42.3	57.6	0.0	0.0
SA1 – Complete cases	19	24	QALYs (Range: 0 - 1)	-1575 (-7951 to 4800)	-0.005 (-0.89 to 0.079)	310 000	5.2	41.9	29.7	23.3
	19	24	Surgery (Range: 0-1)	1727 (-4786 to 8241)	-0.518 (-0.768 to -0.267)	-3334	70.0	30.0	0.0	0.0
SA2 – Human capital	27	29	QALYs (Range: 0 - 1)	-2883 (-9182 to 2696)	-0.004 (-0.051 to 0.043)	736 761	2.0	42.9	40.0	14.9
approach	27	29	Surgery (Range: 0-1)	-272 (-7178 to 4925)	-0.473 (-0.207 to 0.738)	576	48.5	51.5	0.0	0.0
Second main	27	29	QALYs (Range: 0 - 1)	-1918 (-4801 to 1289)	-0.001 (-0.047 to 0.045)	2 918 245	2.7	46.7	42.9	7.7
analysis – Healthcare	27	29	Surgery (Range: 0-1)	809 (-1671 to 2872)	-0.473 (-0.737 to 0.208)	-1713	76.0	24.0	0.0	0.0
perspective										
Analyses above were adjusted for age, sex, depressive feelings, anxiety feelings, education level, living together, nationality, medication, smoking, pain catastrophizing, employment, patients expectations, leg pain, back pain, comorbidities, duration complaints, sick days, work presentation										

Figure 2a: Cost-effectiveness plane for combination therapy intervention for sciatica patients on the waiting list for surgery compared to no intervention (QALY)

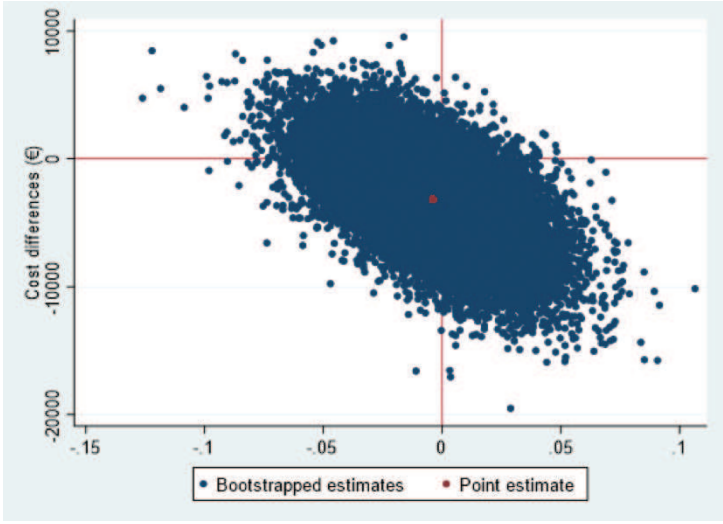


Figure 2b: Cost-effectiveness analysis for combination therapy intervention for sciatica patients on the waiting list for surgery compared to no intervention (Surgery prevented)

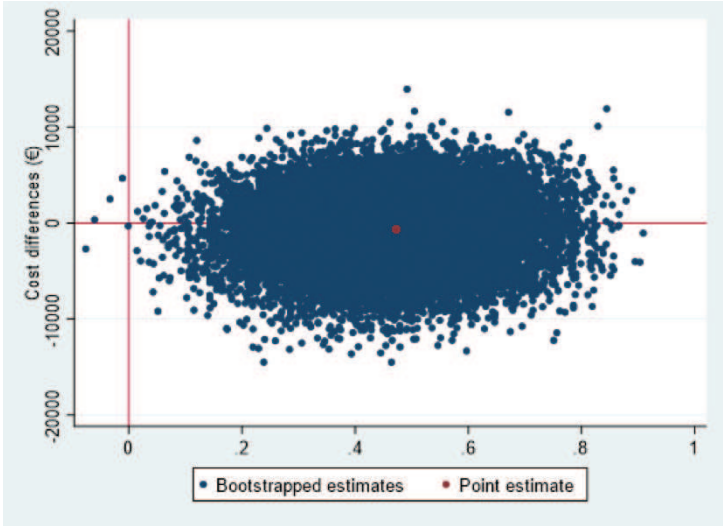


Figure 3a: QALY Cost-effectiveness acceptability curve for combination therapy (intervention) for sciatica patients on the waiting list for surgery compared to usual care

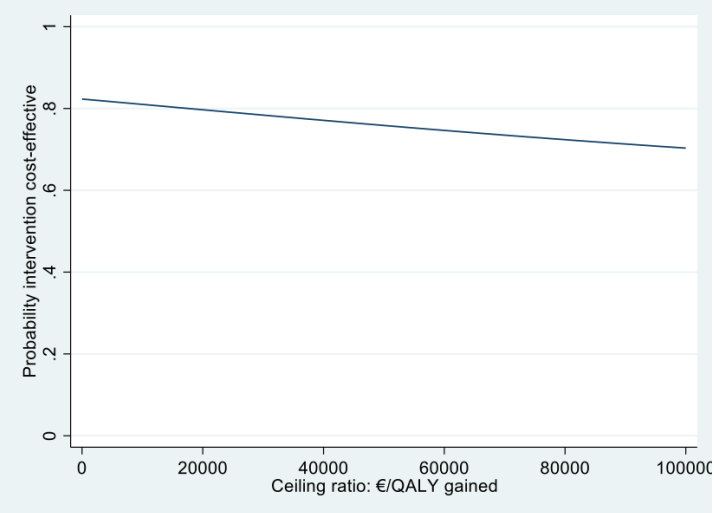


Figure 3b: Cost-effectiveness acceptability curve for combination therapy intervention for sciatica patients on the waiting list for surgery compared to usual care per surgery prevented

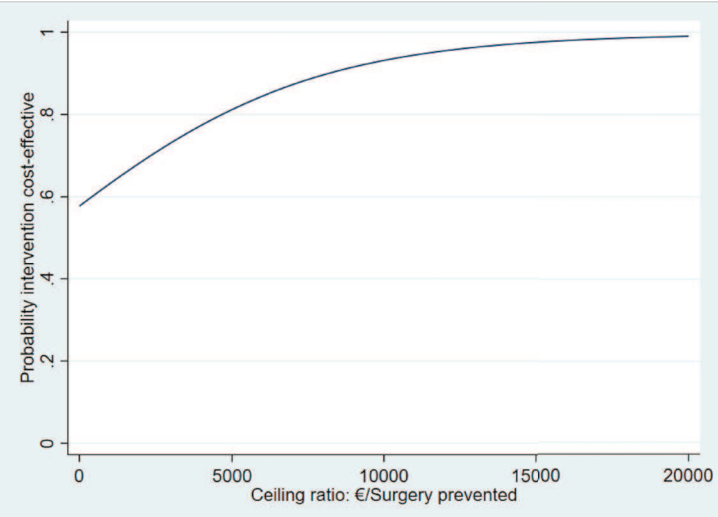
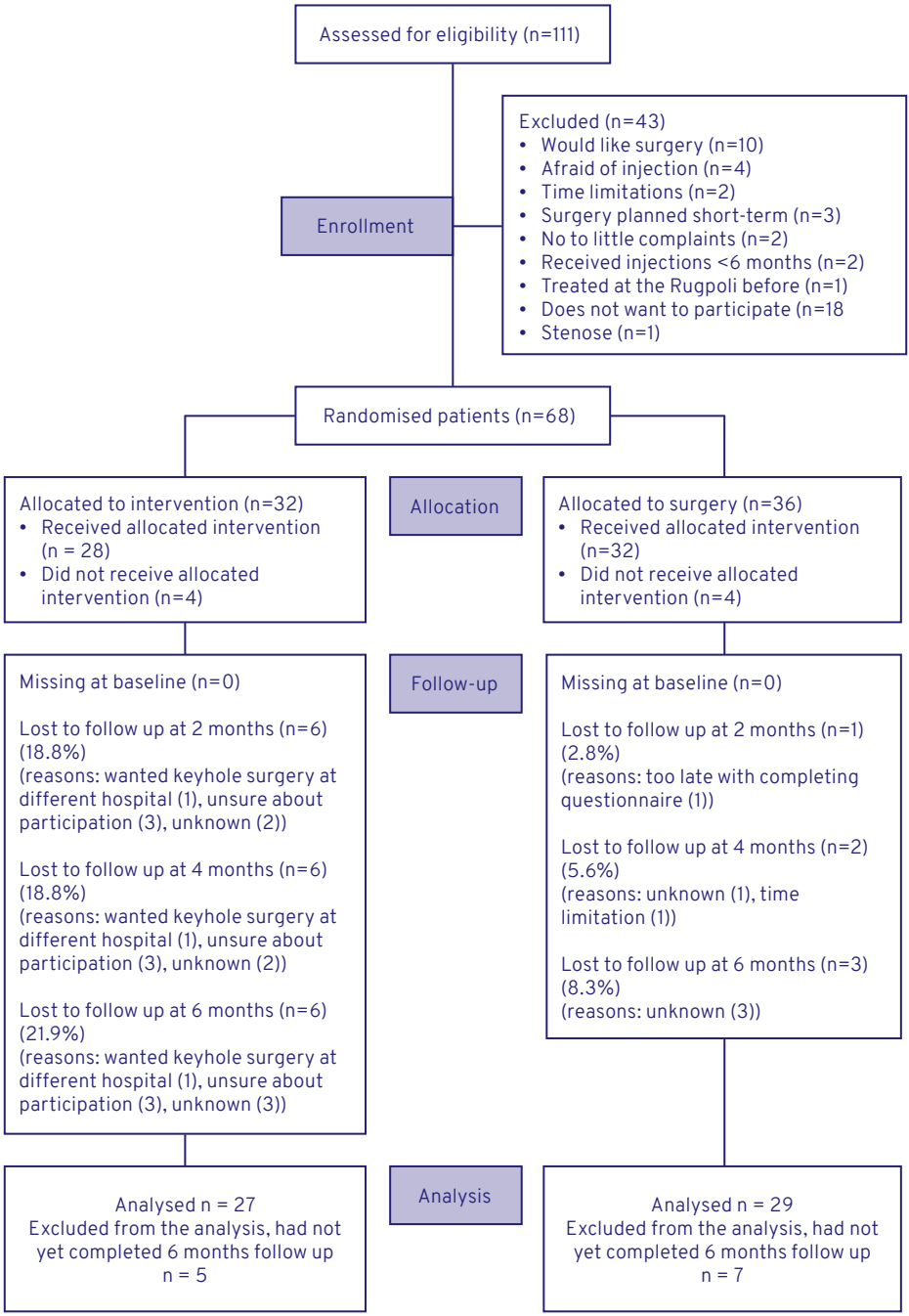


Figure 1: Flow diagram of patients entering the study



REFERENCES

1. Longo DL, Zafonte RD. Sciatica. *N Engl J Med*. 2015;372:1240-8.

2. Oosterhuis T, Costa LOP, Maher CG, de Vet HC, van Tulder MW, Ostelo RW. Rehabilitation after lumbar disc surgery. *Cochrane Database of Systematic Reviews*. 2014;3.

3. Stafford MA, Peng P, Hill DA. Sciatica: a review of history, epidemiology, pathogenesis, and the role of epidural steroid injection in management. *British Journal of Anaesthesia*. 2007;99(4): 461-73.

4. Porchet F, Wietlisbach V, Burnand B, Daeppen K, Villemure JG, Vader JP. Relationship between severity of lumbar disc disease and disability scores in sciatica patients. *Neurosurgery*. 2002;50(6):1253-60.

5. Heliövaara M, Impivaara O, Sievers K, Melkas T, Knekt P, Korpi J, Aromaa A. Lumbar disc syndrome in Finland. *J Epidemiol Community Health*. 1987 Sep;41(3):251-8.

6. Konstantinou K, Hider SL, Jordan JL, Lewis M, Dunn KM, Hay EM. The impact of low back-related leg pain on outcomes as compared with low back pain alone: a systematic review of the literature. *Clin J Pain*. 2013 Jul;29(7):644-54.

7. Schaafstra A, Spinnewijn W, Bons S, Borg M, Koes B, Ostelo R, Spijker-Huiges A, Burgers J, Bouma M, Verburg A. NHG-Standaard Lumbosacraal radiculair syndroom (Tweede herziening). *Huisarts Wetenschap*. 2015;58(6):308-20.

8. Spijker-Huiges A, Groenhof F, Winters JC, van Wijhe M, Groenier KH, van der Meer K. Radiating low back pain in general practice: incidence, prevalence, diagnosis, and long-term clinical course of illness. *Scand J Prim Health Care*. 2015;33(1):27-32.

9. Nielen MMJ, Hek K, Schermer TRJ. Incidentie en prevalentie

van gezondheidsproblemen in de Nederlandse huisartsenpraktijk in 2017. NIVEL Zorgregistraties eerste lijn [internet, [Incidence and prevelance of health problems in the Dutch general practice in 2017]. Incidentie en prevalentie van gezondheidsproblemen in de Nederlandse huisartsenpraktijk in 2017. 2019. Accessed 11-11-2019.

10. Cherkin DC, Deyo RA, Loeser JD, Bush T, Waddell G. An international comparison of back surgery rates. *Spine*. 1994;19:1201-6.

11. Koes BW, van Tulder MW, Peul WC. Diagnosis and treatment of sciatica. *BMJ*. 2007;334(7607):1313–7.

12. Schaafstra A, Spinnewijn WEM, Bons S, Borg M, Koes BW, Ostelo R, Spijker-Huiges A, Burgers JS, Bouma M, Verburg AFE. NHG-Standaard Lumbosacraal radiculair syndroom (tweede herziening). *Huisarts & Wetenschap*. 2015;58(6):308-20

13. Stochkendahl MJ, Hartvigsen J, Kongsted A, Aaboe J, Andersen M, Andersen MØ, Fournier G, Højgaard B, Jensen MB, Jensen LD, Karbo T, Kirkeskov L, Melbye M, Morsel-Carlsen L, Nordsteen J, Palsson TS, Rasti Z, Silbye PF, Steiness MZ, Tarp S and Vaagholt M. National Clinical Guidelines for non-surgical treatment of patients with recent onset low back pain or lumbar radiculopathy. *Eur Spine J*. 2018 Jan;27(1):60-75.

14. Konstantinou K, Dunn KM. Sciatica Review of epidemiological studies and prevalence estimates. *Spine*. 2008;33(22):2464-72.

15. Weinstein JN, Lurie JD, Olson PR, Bronner KK, Fisher ES. United States' trends and regional variations in lumbar spine surgery:1992-2003. *Spine*. 2006;31(23):2707-14.

16. Nederlandse Zorgautoriteit. Open data van de Nederlandse Zorgautoriteit. 2014-2020. https://www.opendisdata.nl/. Accessed 18-03-2020 2020.

17. Arts MP, Brand R, van den Akker ME, Koes BW, Bartels RH, Peul WC, Leiden-The Hague Spine Intervention Prognostic

Study Group (SIPS). Tubular diskectomy vs conventional microdiskectomy for sciatica: a randomized controlled trial. *Jama*. 2009;302(2):149-58.

18. Arts MP, Brand R, van den Akker ME, Koes BW, Bartels RH, Tan WF, Peul WC. Tubular diskectomy vs conventional microdiskectomy for the treatment of lumbar disk herniation: 2-year results of a double-blind randomized controlled trial. *Neurosurgery*. 2011 Jul;69(1):135-44; discussion 144.

19. Arts MP, Peul WC, Leiden–The Hague Spine Intervention Prognostic Study Group (SIPS). Timing and minimal access surgery for sciatica: a summary of two randomized trials. *Acta Neurochir*. 2011.

20. Arts MP, Brand R, van den Akker ME, Koes BW, Bartels RH, Tan WF, Peul WC. Tubular diskectomy vs conventional microdiskectomy for the treatment of lumbar disk herniation: 2-year results of a double-blind randomized controlled trial. *Neurosurgery*. 2011 Jul;69(1):135-44; discussion 144.

21. Clark R, Weber RP, Kahwati L. Surgical Management of Lumbar Radiculopathy: a Systematic Review. *Journal of general internal medicine*. 2019:1-10.

22. Bailey CS, Rasoulinejad P, Taylor D, Sequeira K, Miller T, Watson J, Rosedale R, Bailey SI, Gurr KR, Siddiqi F, Glennie A, Urquhart JC. A. Surgery versus conservative care for persistent sciatica lasting 4 to 12 months. *New England Journal of Medicine*. 2020;382(12):1093-102.

23. Peul WC, van Houwelingen HC, van den Hout, WB, Brand R, Eekhof JA,Trans JT, Thomeer RT, Koes BW. Leiden-The Hague Spine Intervention Prognostic Study Group Surgery versus Prolonged Conservative Treatment for Sciatica. *The New England Journal of Medicine*. 2007;356(22):2245-56.

24. Manchikanti L, Falco, FJ, Singh V, Pampati V, Parr AT, Benyamin RM, ... & Hirsch, J. A. Utilization of interventional techniques in managing chronic pain in the Medicare population:

analysis of growth patterns from 2000 to 2011. *Pain Physician*. 2012;15(6):E969-E82.

25. Oliveira CB, Maher CG, Ferreira ML, Hancock MJ, Oliveira VC, McLachlan AJ, Koes BW, Ferreira PH, Cohen SP, Pinto RZ. Epidural corticosteroid injections for lumbosacral radicular pain. *Cochrane Database of Systematic Reviews*. 2020;4.

26. Chou R, Hashimoto R, Friedly J, Fu R, Bougatsos C, Dana T, Sullivan SD, Jarvik J. Epidural corticosteroid injections for radiculopathy and spinal stenosis: a systematic review and meta-analysis. *Annals of internal medicine*. 2015;163(5):373-81.

27. Pinto RZ, Maher CG, Ferreira ML, Hancock M, Oliveira VC, McLachlan AJ, Koes B, Ferreira PH. Epidural corticosteroid injections in the management of sciatica: a systematic review and meta-analysis. *Ann Intern Med*. 2012 Dec 18;157(12):865-77.

28. van Helvoirt H, Apeldoorn AT, Ostelo RW, Knol DL, Arts MP, Kamper SJ, van Tulder MW. Transforaminal epidural steroid injections followed by mechanical diagnosis and therapy to prevent surgery for lumbar disc herniation. *Pain Med*. 2014;15(7):1100-8.

29. Schenk R, Bhaidani T, Melissa B, Kelley J, Kruchowsky T. Inclusion of Mechanical Diagnosis and Therapy (MDT) in the Management of Cervical Radiculopathy: A Case Report. *The Journal of manual & manipulative therapy*. 2008;16(1).

30. Kilpikoski S, Alèn, M, Paatelma M, Simonen R, Heinonen A, Videman T. Outcome comparison among working adults with centralizing low back pain: Secondary analysis of a randomized controlled trial with 1-year follow-up. *Advances in Physiotherapy*. 2009;11(4):210-7.

31. Mutubuki EN, van Helvoirt, H, van Dongen JM, Vleggeert Lankamp C L, Huygen FJ, van Tulder MW, Klopper-Kaas AH, Ostelo RW. Cost effectiveness of combination therapy (Mechanical Diagnosis and Treatment and Transforaminal Epidural Steroid Injections) among patients with an indication for a Lumbar

Herniated Disc surgery: Protocol of a randomized controlled trial. *Physiotherapy Research International*,. 2019;25(1).

32. Mutubuki EN, Luitjens MA, Maas ET, JPM Huygen, F., Ostelo, R. W., van Tulder, M. W., & van Dongen, J. M. . Predictive Factors of High Societal Costs among Chronic Low Back Pain Patients. *European Journal of Pain*. 2019.

33. Mutubuki EN, van Helvoirt, H, van Dongen JM, Vleggeert Lankamp C L, Huygen FJ, van Tulder MW, Klopper-Kaas AH, Ostelo RW. Cost effectiveness of combination therapy (Mechanical Diagnosis and Treatment and Transforaminal Epidural Steroid Injections) among patients with an indication for a Lumbar Herniated Disc surgery: Protocol of a randomized controlled trial. *Physiotherapy Research International*,. 2019;25(1).

34. Chiarotto A, Ostelo RW, Turk DC, Buchbinder R, Boers M. Core outcome sets for research and clinical practice. *Brazilian Journal of Physical Therapy*. 2017;21(2):77-84.

35. Williamson A, Hoggart B. Pain: A review of three commonly used pain rating scales. *J Clin Nurs*. 2005;14(7):798-804.

36. Patrick DL, Deyo RA, Atlas SJ, Singer DE, Chapin A, Keller RB. Assessing health-related quality of life in patients with sciatica. *Spine* (Phila Pa 1976). 1995 1;20(17):1899-909.

37. Kent P, Grotle M, Dunn KM, Alber HB, Lauridsen HH. Rasch Analysis of the 23-item version of the Ronald Morris disability questionnaire. *Journal of Rehabilitation Medicine*. 2015;47(4):356-64.

38. Yamato TP, Maher, CG, Saragiotto BT, Catley MJ, McAuley JH. The Roland–Morris Disability Questionnaire: one or more dimensions? *European Spine Journal*. 2017;26(2):301-8.

39. Versteegh MM, Vermeulen KM, Evers SM, de Wit GA, Prenger R, Stolk E. A. Dutch Tariff for the five level version of EQ 5D. *Value in Health*. 2016;19:343–52.

40. Versteegh MM, Vermeulen KM, Evers SM, de Wit GA, Prenger R, Stolk EA. Dutch Tariff for the Five-Level Version of EQ-5D. *Value in Health*. 2016;19:343-52.

41. Drummond MF, Schulpher MJ, Torrance GW, O'Brien J, Stoddart, G.L. Oxford: Oxford University Press, 2005. Methodsfor the economic evaluation of health care programmes. New York: Oxford University Press; 2005.

42. Netherlands Bureau voor de Statistiek. Jaarmutatie consumentenprijsindex; vanaf 1963. Central Bureau voor de Statistiek, Den Haag, the Netherlands. https://opendata.cbs.nl/#/CBS/nl/dataset/70936ned/table. Accessed 9 June 2020 2020.

43. Hakkaart-van Roijen L, Van der Linden N, Bouwmans C, Kanters T, Tan S. Costing manual: Methodology of costing research and reference prices for economic evaluations in healthcare. Diemen, the Netherlands: Zorginstituut Nederland. 2015.

44. Koopmanschap MA. PRODISQ: a modular questionnaire on productivity and disease for economic evaluation studies. *Expert Rev Pharmacoecon Outcomes Res*. 2005;5(1):23-8.

45. Koopmanschap MA, Rutten FF, van Ineveld BM, van Roijen L. The friction cost method for measuring indirect costs of disease. *J Health Econ*. 1995 Jun;14(2):171-89.

46. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Statist Med*. 2011;30:377-99.

47. Faria R, Gomes M, Epstein D, White IR. A guide to handling missing data in cost-effectiveness analysis conducted within randomised controlled trials. *Pharmacoeconomics*. 2014;32(12):1157-70.

48. Bhaskaran K, Smeeth L. What is the difference between missing completely at random and missing at random?

International journal of epidemiology. 2014;43(4):1336–9.

49. White IR, Royston P, Wood AM Multiple imputation using chained equations: issues and guidance for practice. *Stat Med*. 2011;30(4):377–99.

50. Zhang J, Kai FY. What's the relative risk?: A method of correcting the odds ratio in cohort studies of common outcomes. *Jama*. 1998;280(19):1690-1.

51. Twisk J, de Boer M, de Vente W, Heymans M. Multiple imputation of missing values was not necessary before performing a longitudinal mixed-model analysis. *Journal of clinical epidemiology*. 2013;66(9):1022-8.

52. Manca A, Hawkins N, Sculpher MJ. Estimating mean QALYs in trial-based cost-effectiveness analysis: the importance of controlling for baseline utility. *Health Econ*. 2005;14(5):487-96.

53. Black WC. The CE plane: a graphic representation of cost-effectiveness. *Medical Decision Making*. 1990;10(3).

54. Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart G.L. Methods for the Economic Evaluation of Health Care Programmes. New York: Oxford University Press; 2005.

55. Fenwick E, O'Brien BJ, Briggs A. Cost-effectiveness acceptability curves-facts fallacies and frequently asked questions. *Health Economics*. 2004;13(5):405-15.

56. Kamper SJ, Ostelo RW, Knol DL, Maher CG, de Vet HC, Hancock MJ. Global Perceived Effect scales provided reliable assessments of health transition in people with musculoskeletal disorders, but ratings are strongly influenced by current status. *Journal of clinical epidemiology*. 2010;63(7):760-6.

57. Bhatia A, Flamer D, Shah PS, Cohen SP. Transforaminal epidural steroid injections for treating lumbosacral radicular pain from herniated intervertebral discs: a systematic review and meta-analysis. *Anesthesia & Analgesia*. 2016;122(3):857-70.

58. Pennington Z, Swanson MA, Lubelski D, Mehta V, Alvin MD, Fuhrman H, Mroz TE. Comparing the Short-Term Cost-Effectiveness of Epidural Steroid Injections and Medical Management Alone for Discogenic Lumbar Radiculopathy. *Clinical Neurology and Neurosurgery*. 2020.

59. Jönsson B. Ten arguments for a societal perspective in the economic evaluation of medical innovations. *Eur J Health Econ*. 2009;10:357–9.

60. Brouwer WB, Van Exel NJA, Baltussen RM, Rutten FF. A dollar is a dollar is a dollar—or is it? *Value in Health*. 2006;9(5):341-7.

61. Willan AR, Briggs AH, Hoch JS. Regression methods for covariate adjustment and subgroup analysis for non-censored cost-effectiveness data. *Health Econ*. 2004;13(5):461-75. doi:10.1002/hec.843.

62. Willan AR. Statistical analysis of cost-effectiveness data from randomized clinical trials. *Expert Rev Pharmacoecon Outcomes Res*. 2006 Jun;6(3):337-46.

63. Gabrio A, Mason AJ, Baio G. Handling missing data in within-trial cost-effectiveness analysis: a review with future recommendations. PharmacoEconomics-open. 2017;1(12):79-97.

64. Díaz-Ordaz K, Kenward MG, Cohen A, Coleman CL, Eldridge S. Are missing data adequately handled in cluster randomised trials? A systematic review and guidelines. *Clinical Trials*. 2014;11(5):590-600.

65. Noble SM HW, Tilling K. Missing data in trial based cost effectiveness analysis: the current state of play. *Health economics*. 2012;21:187-200.

66. Doshi JA, Glick HA, Polsky D. Analyses of cost data in economic evaluations conducted alongside randomized controlled trials. *Value Health*. 2006;9(5):334-40.

67. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL,

Torrance GW. Methods for the economic evaluation of health care programmes. Oxford university press; 2015.

68. Nixon RM, Thompson SG. Methods for incorporating covariate adjustment, subgroup analysis and between centre differences into cost effectiveness evaluations. *Health economics*. 2005;14(12):1217-29.

69. Petrou S, Gray A. Economic evaluation alongside randomised controlled trials: design, conduct, analysis, and reporting. *BMJ*. 2011;342(1548).

70. Ramsey S, Willke R, Briggs A, Brown R, Buxton M, Chawla A et al. Good research practices for cost-effectiveness analysis alongside clinical trials: the ISPOR RCT-CEA Task Force report. *Value Health*. 2005;8(5):521-33.

71. Ramsey SD, Willke RJ, Glick H, Reed SD, Augustovski F, Jonsson, B. Sullivan SD. Cost-effectiveness analysis alongside clinical trials II—an ISPOR Good Research Practices Task Force report. *Value in Health*. 2015;18(2):161-72.

72. Nixon RM, Thompson SG. Methods for incorporating covariate adjustment, subgroup analysis and between centre differences into cost effectiveness evaluations. *Health economics*. 2005;14(12):1217-29.

73. van Asselt AD, van Mastrigt GA, Dirksen CD, Arntz, A, Severens JL, Kessels AG. How to deal with cost differences at baseline. *Pharmacoeconomics*. 2009;27(6):519-28.

74. Nederhof AJ. Methods of coping with social desirability bias: A review. *European journal of social psychology*. 1985;15(3):263-80.

6

Exercise therapy for sciatica. Is it effective? A systematic review and meta-analysis

Seiger A.,[1] Mutubuki E.N.,[1] van Tulder M.W.,[1,2] Vleggeert-Lankamp C.L.A.M.,[3] Peul W.C.,[3] van Susante J.L.C.,[4] Harhangi B.S.,[5] Ostelo R.W.J.G,[1, 6] Jansma E.P.,[7] Rubinstein S.M.[1]

[1] Department of Health Sciences, Faculty of Science, Amsterdam Movement Science Research Institute, Vrije Universiteit Amsterdam, The Netherlands.
[2] Department of Physiotherapy & Occupational Therapy, Aarhus University Hospital, Aarhus, Denmark.
[3] Department of Neurosurgery, Leiden University Medical Centre, LUMC, The Netherlands.
[4] Department of Orthopedics, Rijnstate Hospital, The Netherlands.
[5] Department of Neurosurgery, Erasmus MC: University Medical Centre Rotterdam, The Netherlands.
[6] Department of Epidemiology and Biostatistics, Amsterdam Public Health Research Institute, VU University Medical Center, Amsterdam, the Netherlands.
[7] Medical Library, VU University Medical Centre, Amsterdam, Netherlands.

Under Review

ABSTRACT

Purpose: A systematic review was performed to determine the effectiveness of exercise therapy on sciatica.
Method: A comprehensive search was conducted in several electronic databases (inception - May 2019) to identify relevant randomized controlled trials. Exercise was compared to 1) other therapies, 2) surgery 3) no therapy, 4) sham exercises, and 5) exercise as an adjunct therapy. Mean difference (MD), standardized mean difference (SMD), relative risk (RR), absolute risk difference (RD), and the number to treat (NNT) were calculated. Primary outcomes were pain, functional status, and global perceived effect at short, intermediate, and long-term follow-up. Random effect models were used for meta-analyses.
Results: 9 RCTs (997 participants) were included. Exercise had a small short-term effect on functional status compared to other therapies [MD -8.4 (95% CI -15.70, -1.10); 0-100 scale]; large short-term effect on pain compared to no therapy [MD -2.07 (95% CI -3.24, -0.89), 0-10 scale]; and a medium long-term effect on global perceived effect when given as an adjunct therapy [RR 1.42 (95% CI 1.11, 1.81); RD 0.23; NNT=4]. Other comparisons had small, non-statistically significant differences. Certainty of evidence was low to very low for all outcomes, and statistical heterogeneity could not be explained. None of the studies compared exercises to surgery.
Conclusion: Based on this review there is no consistent effect of exercise for the treatment of sciatica. However, better quality studies are highly recommended. Prospero registration [CRD42018080659].

Keywords: Systematic review, meta analyses, exercise, sciatica, herniated disc.

BACKGROUND

Sciatica, also referred to lumbosacral radicular syndrome or radiculopathy, is a common condition. It has a life time prevalence which varies from 12.2% to 43% and has a point prevalence ranging from 1.6% to 13.4% [1-6]. Sciatica is mainly characterized by radiating pain in the leg and is sometimes associated with sensory, motor or reflex deficits [7]. Symptoms of sciatica result from a deficit or irritation of a particular lumbar or sacral nerve root [7]. Sciatica is associated with increased pain, increased disability, and lower quality of life [8-10].

A wide range of different therapies including surgery medication, exercises, and injections are used to treat sciatica[10]. Exercises are often given under the supervision of a therapist, [11-13]. In clinical settings, exercises are mostly prescribed to reduce pain, improve physical functioning, or speed-up recovery [14]. Exercise is associated with several health benefits, including improvement of the cardiorespiratory and cardiovascular system, neuro-coordination, postural musculature, muscle strength, and stabilization [14,15]. However, there is no international consensus regarding the benefits of exercise for patients with sciatica. The UK NICE guideline [NG59] recommends a group exercise program (e.g. aerobic, mind–body or a combination of approaches), while the Dutch general practitioner guideline discourages routine referrals for exercise therapy and recommends limiting activities of daily living (ADL) [16,4].

Systematic reviews on exercise therapy for sciatica patients are either outdated, have specific comparisons, or do not focus on sciatica, exclusively [12,17,13].

METHODS

The objective of this systematic review is to determine the effectiveness of exercise therapy in patients with sciatica. We expected exercise to be similar compared to other therapies and surgery; superior compared to no therapy and sham exercises; inferior compared to exercise as an adjunct therapy. This systematic review is registered in PROSPERO, register number [CRD42018080659]. A more detailed description of the methods can be found in the protocol, which can be uploaded from the PROSPERO database.

Deviation from the protocol

This review deviates from the protocol on the following issues: We originally had 2 comparisons (1) exercise compared to other therapies and 2) exercise compared to surgery), but ultimately we decided to add the comparisons: 3) Exercise compared to no therapy; 4) exercise compared to sham exercises; and 5) exercise as an adjunct therapy. Comparisons were added in order to reduce clinical heterogeneity and improve interpretation of the effect of exercises.
We intended to report leg and back pain, separately. However, one study reported leg pain only and one study reported both leg and back pain [18,19]. Other studies did not mention whether back or leg pain was measured, therefore, we combined all studies on back and leg pain.
Originally, trials with a mixed population of patients (e.g. patients with and without sciatica) would be excluded, but we changed this into: studies were included if at least 70% of the patients had sciatica.
Finally, clinical relevance assessment of dichotomous outcomes was not defined in the protocol. We used the recommendations of the Cochrane Back and Neck group [20].

Criteria for considering studies for this review

Types of studies

Only randomized controlled trials (RCTs) were included.

Observational and uncontrolled studies were excluded, as were pseudo-randomized controlled trials (e.g. studies where treatment allocation was not blinded).

Types of participants

Sciatica is characterized by the symptom of radiating leg pain caused by the compression of the sciatic nerve or a compression of the nerve root that will pass into the sciatic nerve. Studies involving adult participants with sciatica or derivations of sciatica, e.g. lumbosacral radicular syndrome or radiculopathy were included.

Type of intervention

Trials studying any type of exercises were included. Studies on exercises in combination with other therapies were excluded, unless the effect of exercises could be extracted (exercise as an adjunct therapy). Exercise was defined as follows: "a series of specific movements with the aim of training or developing the body by a routine practice or as physical training to promote good physical health" [21].

Types of comparisons

Exercise was compared to 1) other therapies, 2) surgery, 3) no therapy, 4) sham exercises, and 5) exercise as an adjunct therapy (i.e. included if the contrast reflected the effect of exercises).

Types of outcome measures

The primary outcomes are:

- Pain (leg and/or back pain);
- Functional status;
- Global perceived effect.

The secondary outcomes are:

- Quality of life;
- Vocational outcomes;
- Physical outcomes.

Search methods for identification of studies

The search was performed in the following databases: PUBMED, EMBASE, Physiotherapy Evidence Database (PEDro), CINAHL, and the Cochrane Library, from the inception of the database

to May 2019. Reference lists of existing systematic and included studies were examined to identify possible missed RCTs. There were no restrictions on language or publication status. Appendix I contains the full electronic data search.

Data collection and analysis

We followed the methods recommended by the Cochrane Back and Neck group [22]. The selection of studies, risk of bias assessment, and data extraction was independently performed by 2 or 3 authors (AS, ENM, SMR). Disagreements were resolved through discussion or if necessary, a third reviewer was included as an arbiter (SMR, MWvT or RWJGO). The evaluation of the certainty of the evidence was performed by AS and was checked by ENM or SMR.

Selection of studies

Only full papers were included. Abstracts and proceedings from congresses or any other 'grey literature' were excluded. Each included article is listed in table 1 and the selection process may be found in figure 1. The authors of this review were not blinded to the authors of the papers nor to the journal of publication. Authors were contacted if clarification of the results was necessary.

Assessment of risk of bias of included studies

We assessed the following domains: selection bias, performance bias, detection bias, attrition bias, selective outcome reporting bias, as well as any other bias (e.g. publication bias) (figure 2). The operational definitions recommended by the Cochrane Back and Neck group were used (appendix II) [22].

Measures of treatment effect

Pooled effects for continuous outcomes were measured using a mean difference (MD) with 95% confidence intervals (CI) or a standardized mean difference (SMD) when different scales were used. SMD and 95% CI values were back-transformed multiplying by a representative standard deviation (SD). SMD was back-transformed into a 0-10 scale for pain and into a 0-24 scale for functional status (the higher the worse). For dichotomous outcomes, a risk ratio (RR) with 95% CI and absolute risk

difference (RD) are presented. The number needed to treat was calculated in order to determine the size of the effect for harm and benefits. All effects were stratified by the follow-up period: short- intermediate-, and long-term (closest to 3, 6 and 12 months, respectively). Continuous outcomes were transformed to the same direction; a negative effect size means a larger effect for exercise therapy. In case of a cross-over, trial data were extracted before the cross-over [23]. In factorial design RCTs, exercise was given as individual or combination therapy (with either traction, manipulation, a corset or a combination of these treatments). To distill the effects of exercise, we only used data where exercise was compared to no exercise. Effects were estimated using the generic inverse-variance method and a random effects model was used for all analyses. We used Review Manager 5.3.

Assessment of clinical relevance

The clinical relevance for the pain, functional status and global perceived effect scores were determined based on the recommendations of the Cochrane Back and Neck group [20]:

- Small clinical relevance: MD < 10%; SMD < 0.5; RR <1.25 or > 0.8.
- Medium clinical relevance: MD 10%- 20%; SMD 0.5 - 0.8; RR 1.25 - 2.0 or 0.5 - 0.8.
- Large clinical relevance: MD ≥ 20% ; SMD ≥ 0.8; RR >2.0 or < 0.5.

Dealing with missing data

We tried to contact the authors of the included studies when data were missing or incomplete. When the median was reported instead of the mean and sample size was large (>70), we assumed that the median was similar to the mean and that the width of the of interquartile range (IQR) would be 1.35 times the SD (Higgins, Deeks [24], section 7.7.3.5). If, baseline and final SD were presented, but change SD were missing, we used calculations recommended by the Cochrane handbook to estimate SD (Higgins et al. [25], section 16.1.3.2).

Assessment of heterogeneity

Statistical heterogeneity was assessed by visual inspection of

the forest plots, the Q-test, and the I^2 statistic [26].
In cases of substantial heterogeneity, we explored the data further, and compared the characteristics of individual studies and conducted subgroup analyses, where possible. Statistical heterogeneity was classified according to Cochrane handbook [26] appendix III.

Assessment of reporting biases
Because each comparison consisted of a low number of studies, funnel plots for examining publication bias were not plotted. Protocols were compared to examine possible selective outcome reporting bias.

Data synthesis
GRADE was used to summarize the certainty and strength of the evidence. GRADE was based upon the following determinants: limitations of design, imprecision, inconsistency, indirectness, and other factors (e.g. publication bias). We used the definitions of the Cochrane Back and Neck group [22,27], appendix IV.

Subgroup analysis and investigation of heterogeneity
If data permitted, the following subgroup analyses were performed:
- Clinical condition (e.g. MRI-confirmed sciatica);
- Characteristics of exercise (e.g. strengthening versus mobilization types of exercise or type of delivery: individual exercises versus group exercises);
- Different types of other therapies.

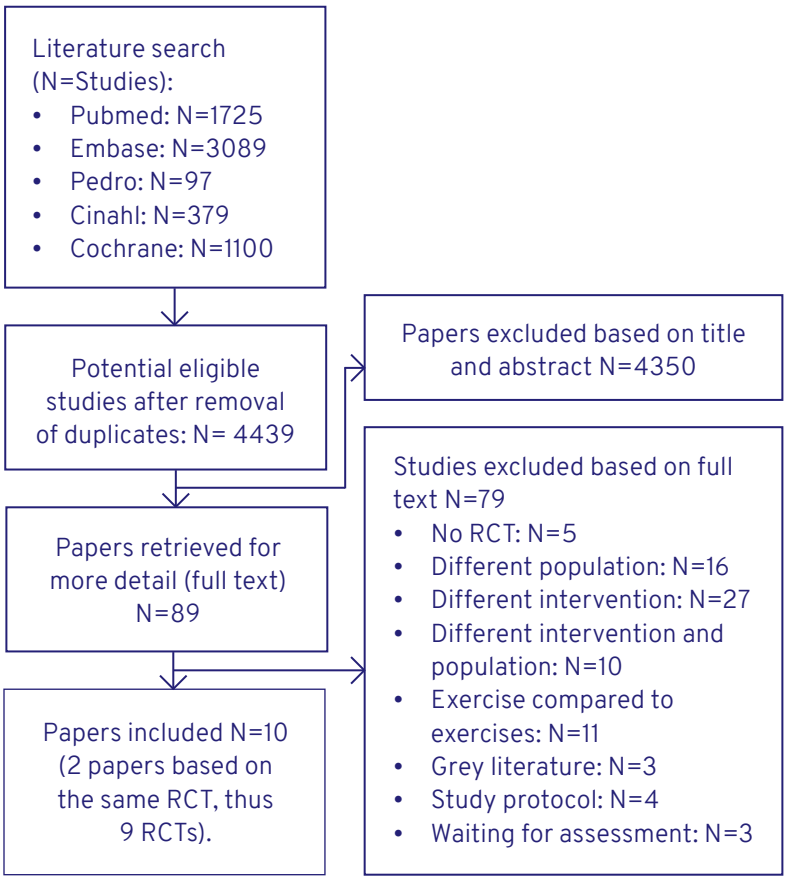
Sensitivity analysis
Sensitivity analyses conducted, were thought to be clinically meaningful, in order to determine the robustness of the data and in order to explain possible sources of heterogeneity between studies. The following sensitivity analyses were conducted:
- Assumptions surrounding missing data for the meta-analyses.
- Only concealed treatment allocation.

RESULTS

Results of the search

In total, 4439 papers were screened on title and abstract, 89 papers were retrieved for full text, and 9 RCTs represented in 10 papers were included (Figure 1). The study of Verwoerd et al. [28] was a subgroup analysis from the RCT of Luijsterburg et al. [19]. No additional RCTs were identified through reference checking. We were able to assess papers written in other languages than English, except for 2 Chinese articles [29,30]. 1 article form Indonesia was excluded because we could not retrieve this article [31]. Reasons for exclusion are presented in figure 1.

Figure 1: Flow chart of inclusion and exclusion systematic review.



Description of included studies

Type of population: The 9 RCTs included 997 patients with sciatica, lumbar herniated disc, lumbar radicular syndrome, lumbar posterior derangement with neurological signs, or radiating pain corresponding to a lesion at L5-S1. Clinical diagnosis was confirmed by CT, MRI or myelography in 4 studies [23,32-34]. Average age ranged from 33 [23] to 46 years [18] (range 18 to 76 year). The ratio of men to women was unclear in one study [23]. When reported, there were no relevant differences between the randomized groups for gender [18,33,19]. Duration of symptoms largely differed and ranged from 3 days to more than 10 years [32]. Three studies had a small sample size ≤ 50.[33,35,34] Sample sizes varied from 25 [35] to 322 [36]. The follow-up ranged from 5 days [33] to 12 months [32,19] (Table 1).

Type of intervention: Exercises were given for a short period of time (i.e. minimum 5-7 days and maximum 8 weeks) to improve strength and/or range of motion of the lower back, glutes, or leg muscles (e.g. core stability training, McKenzie, isometric contractions and hold-relax techniques). Treatments were given as daily homework or by visiting a specialist, such as a physical therapist.

Types of comparisons: Exercise was compared to other therapies (N=3 studies) [36,33,35], no therapy (N=2) [23,37], sham exercises (N=1) [18], and exercises given as an adjunct therapy (N=3) [32,19,34]. None of the included studies compared exercises to surgery. Other therapies consisted of manipulation, traction, lumbar support, and joint mobilization, either given as an individual treatment or as a combination [36,33,35]. Adjunct exercises were added to general practitioner care, injections, or acupuncture with massage [32,19,34].

Types of outcome measures: Short-term effect of pain was measured in all 9 RCTs, however, data were not presented in 1 paper [33]. The visual analogue scale (VAS) and numeric rating scale (NRS) were most commonly used (n=6) [18,23,37,19,35,34]. Both outcomes ranged from 0-10 (0 = no pain, 10 = worse pain). Other measurements used include a pain analogue scale (-100 to +100; -100 = worse deterioration of pain ; +100 = most

improvement of pain) and a pain numbness scale (0-20; the higher the better) [36,32]. Two studies reported specifically on leg pain [18,19], while all other studies reported only pain without specifying the region [23,36,32,37,35,34,28].
Short-term effect on functional status was measured in 6 studies [18,32,33,19,35,34], but outcomes were presented in only 5 [18,32,19,35,34]. A large diversity of instruments were used for the functional status and global perceived effect. Two RCTs measured functional status with the Roland Morris Disability questionnaire [(RMDQ), 0-24 scale; the higher the worse (2 RCTs)] [18,19], one with the Oswestry Disability Index [(ODI), 0-100 scale; the higher the worse (1 RCT)] [35], one with the living and working ability scale (0-16 scale; the higher the better) [32], and one with the low back disability questionnaire (DSQI; 0-100 scale; the higher the worse) [34].
Short-term global perceived effect was measured in 6 RCTs, whereas 5 RCTs presented outcomes [36,32,33,19,34]. One paper used the 7-point Likert-scale [19]; other measurements were a simple question (better, worse, or same) [36]; overall clinical assessment (based on pain walking, working and living ability, tenderness, reflexes and straight leg raising) [32]; overall clinical assessment (based on straight leg raising, pain, activities of daily living and mobility of lumbar spine) [33]; and global rating of change [(GROC); -7 to +7 scale; the higher the better] [34]. Dichotomization was already performed by the included studies; some used any improvement as cut-off point, while others used much improvement [19,34].

Drop-outs: Overall the drop-out rate was low (<20%). Only, for the intermediate-term, one study presented a 22% dropout rate at 4 months for the global perceived effect [36].

Side effects and adverse events: Most papers did not report on side effects or adverse events [23,36,32,37,33,19,35]. One paper specifically reported on adverse events, but none were found [34]. One RCT reported rapidly progressing symptoms where referral to the neurosurgeon was necessary (n=6; 3.3%) [18]. Coxhead et al. [36] reported a 7% of drop-out rate across groups because complaints deteriorated during the first 4 weeks.

Table 1. Study characteristics included studies

Study	Methods	Participants	Interventions	Outcomes	Notes
Albert, Manniche [18].	RCT. Measurement moments: • Baseline. • 8 weeks. • 1 year.	Severe sciatica. Patients from secondary care after unsuccessful treatment in primary care. Total participants: N= 181. **I:** 53% females. **C:** 43% females. **I:** 46 years (38-52)[a]. **C:** 44 years (37-51)[a]. Duration of sciatica: Not reported per group: • less than 1 month: 16%. • 1-3 month: 61.3%, • 3-6 months: 17.7%, • 6-12 months: 5%.	**Both groups received:** Information & advice to stay active. Paracetamol and NSAID. 8 weeks of treatment. **I:** Individualized, symptom-guided exercises (standardized): • Exercises based on McKenzie. • Stabilizing exercises for the transverse abdominis and multifidus muscles. • Dynamic exercises for abdominal wall and back extensors. • Multi-disciplinary team: 3 physiotherapists and 1 chiropractor. **C:** Non-individualized, general sham exercises: • Not back related, low-dose exercises. • Multi-disciplinary performed by 2 physiotherapists and 2 chiropractors.	**Pain** (NRS; 0–10 scale; no pain- worst pain; leg pain): **I:** • Baseline: 4.3 ± 2.3. • 8 weeks: 1.5 ± 2.1. • 1 year: 1.5 ± 2.1. **C:** • Baseline: 4.5 ± 2.5. • 8 weeks: 2.3 ± 2.7. • 1 year: 1.4 ± 2.4. **Functional status** (RMDQ; 0-24 scale; no disability-severe disability): **I:** • Baseline: 15.5 (11-18)[a]. • 8 weeks: 6 (2-12)[a]. • 1 year: 3.5 (1-10)[a]. **C:** • Baseline: 15 (12-18)[a]. • 8 weeks: 6 (2-12)[a]. • 1 year: 3.5 (1-10)[a]. **Global perceived effect** (0-5 Likert scale) • Effects not reported per group.	**Dropouts:** 1.1%. **Side effects:** Rapidly progressing symptoms, therefore, patients were referred to the neurosurgical department: N=6 (3.3%). **Adverse events:** Not reported.

Continued - Table 1. Study characteristics included studies

Bakhtiary et al. [23].	RCT: Cross over trial. After 4 weeks cross over. Measurement moments: • Baseline. • 4 weeks. • 8 weeks. (cross over; could not be used for this review).	CT or MRI confirmed herniated lumbar disc at L4-5 or L5-S1. Patients from outpatient orthopedic clinics. Total participants: N= 60. Gender: not mentioned. Average age: **I:** 33.0 ± 5.1. **C:** 32.6 ± 6.4. Duration sciatica: **I:** 3.5 ± 1.4 months. **C:** 4.4 ± 1.2 months;	**Both groups received:** Exercise, but only first 4 weeks used. Control did not receive any treatment in the first 4 weeks. **I:** • Lumbar stabilizing exercises (core stability exercises). • 4 stages of stabilizing exercises (easy – advanced). • Every week 1 stage for 4 weeks. Performed at home: 2 times a day. 10 times a week. • Accuracy regularly controlled by physiotherapist each week. **C:** No treatment in the first 4 weeks.	**Pain:** VAS (0-10 scale; no pain- worst pain): **I:** • Baseline: 4.29 ± 0.9; • 4 weeks (change scores): -3.2 ± 1.47. **C:** • Baseline: 4.5 ± 1.1. • 4 weeks (change scores): -0.5 ± 1.17. **Functional status:** Not measured. **Global perceived effect:** Not measured.	**Dropouts:** Failed to complete full term of the study: n=8 (13.3%). **Side effects:** Not reported. **Adverse events:** Not reported.

Continued - Table 1. Study characteristics included studies

Coxhead et al. [36].	RCT: Factorial design. 16 groups: 4 different treatments (24). Measurement moments: • Baseline. • 4 weeks. • 4 months. • 16 months.	Patients with pain of sciatic distribution. Pain at least as far as the buttock crease (with or without back pain). Total participants: N= 322 Gender: Not reported per group. Gender: 185 males and 149 females. Age: Not reported per group. Average age: 41.9 ± 12.2 years. Duration of sciatica: Not reported per group. Duration sciatica: 14.3 ± 16.1 weeks.	**Both groups received:** Treatment for 4 weeks. **I:** • Factorial design; Exercises as a single treatment or as a combination treatment to one of the control treatments. • Exercises in all ranges of motion and muscle-groups. **C:** No exercises: Individual or combination treatment of the following 3 control treatments: • Traction. • Manipulation (Maitland). • Corset—lumbar support.	**Pain** (Pain analogue scale (-100 to +100 scale; higher scores mean more improvement): **I:** • Baseline: not presented. • 4 weeks: 49.0 ± 40.0. **C:** • Baseline: not presented. • 4 weeks: 46.3 ± 38.2. **Functional status:** • Results not reported per group. **Global perceived effect** (better, worse or same): Effect: Better. **I:** • 4 weeks: N=120 (80%). • 4 months compared to 4 weeks: N= 85 (69%). **C:** • 4 weeks: N=107 (75%). • 4 months compared to 4 weeks: N= 96 (76%).	Information about type and frequencies of the exercises is limited. **Dropouts:** 4 weeks: 10.6%; 1 months: 6%; 4 months: 22%; 16 months: 20%. **Side effects:** Complaints deteriorated during 4 weeks of treatment period: N= 21 (7%). **Adverse events:** Not reported.

Continued - Table 1. Study characteristics included studies

Ding, Yang [32].	RCT. Measurement moments: • Baseline. • 2-3 weeks. Recurrence rate measured at: • 6 months. • 12 months.	PCT or MRI confirmed herniated disc with radiating pain in lower limbs. Total participants: N= 128 Gender: Not reported per group. Gender: 72 males and 56 females. Age: Not reported per group. Average age: 41 years (range 20-76). Duration of sciatica not reported per group. Duration sciatica: 3 days-10 years (mean 154 days).	**Both groups received:** Treatment: 5 times per week. Total of 2-3 weeks. Electro-acupuncture, low frequency currents, and massage (listed at control section). **I:** Exercise as an adjunct therapy to control. Exercise: • Isometric squatting stances for core muscle training. • Performed once a day until a little sweat comes out. **C:** • Electro-acupuncture on lumbar points with continues waves for 20 minutes. • Low frequency currents for 15 minutes (with a low frequency device). • Massage soft tissue + oblique pulling for 20 minutes.	**Pain** (Pain and numbness; 0-20 scale; higher scores means more improvement): **I:** • Baseline: 8.21 ± 3.16. • 2-3 weeks: 16.07 ± 4.00. **C:** • Baseline: 10.00 ± 3.39. • 2-3 weeks: 14.28 ± 3.85. **Functional status** (Living and working ability (0-16 scale; higher scores means more improvement): **I:** • Baseline: 11.42 ± 3.45. • 2-3 weeks: 15.14 ± 1.70. **C:** • Baseline: 12.06 ± 2.71. • 2-3 weeks: 14.85 ± 1.87. **Global perceived effect** (Overall clinical outcome; 0-100 scale): Effect: Excellent and good (>75 points). **I:** 2-3 weeks: N= 49 (77%). **C:**2-3 weeks: N= 35 (55%).	**I:** Exercises only existed out of 1 isometric exercise. **Dropouts:** Not reported. **Side effects:** Besides recurrence rates, not reported. **Recurrence rate:** **I:** 6 months 7.8%; 1 year 10.9%. **C:** 6 months 17.2%; 1 year 25.0%. **Adverse events:** Not reported.

Continued - Table 1. Study characteristics included studies

Huber et al. [37].	RCT. Measurement moments: • Baseline. • 20 days.	Right-sided sciatica caused by a herniated disc (first episode). Total participants: N= 52. Gender: Not reported per group. Gender: 24 males and 28 females. Age: Not reported per group. Average age: 35 years (range 30-45 years). Duration of sciatica not reported per group. Duration sciatica: Average 14 days.	**Both groups received:** Treatment for 20 days. **I:** • Isometric contractions of trunk extensors and rectus abdominis in a supine lying position. • Isometric contractions of gluteus maximus, quadriceps femoris and extensors digitilongus muscles. Lying position with legs flexed in hips and knees. • Sessions daily: 20 repetitions for each muscle. Average time muscle contraction: 10-s with a resting period at 10-s. **C:** No treatment.	**Pain** (VAS; 0-10 scale; no pain - worst pain): **I:** • Baseline: 7.2 ± 0.9 • After 20 days: 5.2 ± 1.0 **C:** • Baseline: 7.4 ± 0.9 • After 20 days: 6.9 ± 1.3 **Functional status:** Not measured. **Global perceived effect:** Not measured.	Only right sided sciatica. Short-follow up. Small sample size. **Dropouts:** Not reported. **Side effect:** Not reported. **Adverse events:** Not reported.

Continued - Table 1. Study characteristics included studies

Ljunggren et al. [33].	RCT. Measurement moments: • Baseline. • 5-7 days.	Radiating pain, neurological symptoms and signs corresponding to a lesion of the L5 and/or SI nerve root. Lumbar myelogram confirmed indentation of the dural sac and/or widened or shortened nerve root pocket. Total participants: N= 50 **I:** 13 males, 13 females. **C:** 14 males, 10 females. Average age (range 19-62): **I:** 41.1 years. **C:** 42.2 years. Duration sciatica: **I:** 5.3 months. **C:** 4.8 months.	**Both groups received:** Treatment provided by physiotherapists. Individual info + lectures on back care. Crutches and elastic lumbar supports for out-of-bed activities. Analgesics were given on request. Recommendation: 2 hours comfortable lying position after treatment. Treatment period was 5-7 days. **I:** Exercises: • Isometric exercises for abdominal, back, hip and thigh muscles performed in crook, side lying, and in the supine position. • Daily 20 minutes: Muscle contraction for 6-8 seconds. 5-10 times each muscle group. **C:** Manual traction 1 time per day 10 minutes or 2 times per day 5 minutes.	**Pain** (VAS; 0-10 scale; no pain - worst pain): Results were not reported. **Functional status** (activities of daily living graded according Rolland and Morris (1983): Results were not reported. **Global perceived effect** (Conclusion neurologist based on straight leg raising, mobility of lumbar spine, pain, and functional status): Effect: Pain-free or improved. **I:** • 5-7 days: 10 patients (38%) pain free or improved. **C:** • 5-7 days: 10 (42%) patient 'pain free or improved'.	Short-follow up. Small sample size. **Dropouts:** Not reported (n=1 excluded due to improvement). **Side effect:** Not reported. **Adverse events:** Not reported.

Continued - Table 1. Study characteristics included studies

Luijsterburg et al. [19] and post hoc study Verwoerd et al. [28].	RCT. Measurement moments: • Baseline. • 3 weeks. • 6 weeks. • 12 weeks. • 52 weeks.	Primary care patients diagnosed with lumbar radicular syndrome. Total participants: N= 135 **I:** 38 females (57%). **C:** 27 females (40%). Average age: **I:** 42 years ± 10. **C:** 43 years ± 12. Duration sciatica: **I:** 12.1 days ± 10.1. **C:** 14.2 days ± 10.2.	**Both groups received:** Individual treatment. Max 9 treatments in 6 weeks. General practitioner (GP) care. **I:** Exercise therapy as an adjunct therapy to GP care. • Active exercise therapy + info/advice lumbar radicular syndrome. • Session: 30 min **C:** General practitioner care: • Treatment according Dutch clinical guidelines. • Information + advice about lumbar radicular syndrome + pain medication.	**Pain** (NRS 0-10 scale; no pain - unbearable pain; leg pain; average of 3 ratings): **I:** • Baseline: 6.3 ± 2.2. • 6 weeks (change score): -3.0 ± 2.7. • 12 weeks (change score): -3.9 ± 2.8. **C:** • Baseline: 6.3 ± 2.2. • 6 week (change score): -3.3 ± 2.8. • 52 weeks (change score): -3.7 ± 2.7. **Functional status:** RMDQ (0-24 scale; no disability - severe disability): **I:** • Baseline: 15.9 ± 4.1. • 6 weeks (change score): -5.3 ± 7.0. • 12 weeks (change score): -7.7 ± 7.3. **C:** • Baseline: 15.4 ± 5.0. • 6 weeks (change score): -6.6 ± 6.1. • 12 weeks (change score): -8.5 ± 6.7. **Global perceived effect** (1-7 Likert-scale; completely recovered - worse than ever): Effect: completely recovered and much improved. **I:** • 3 weeks: N=30 (45%). • 12 weeks: N=47 (70%). **C:** • 6 weeks: N=30 (44%). • 12 weeks: N=42 (62%).	**Dropouts:** 12 months: <20%. **Drop outs per arm:** **I:** n=7 (10%). **C:** N = 11 (16%). **Side effect:** Not reported. **Adverse events:** Not reported.

Continued - Table 1. Study characteristics included studies

Schenk et al. [35].	RCT. Measurement moments: • Baseline. • End of treatment (After 3rd visit physical therapy).	Patients with lumbar posterior derangement, with or without neurological signs (>75% with neurological signs). Total participants: N= 25 Gender: Not reported per group. 15 females (60%) 20 males (. Age: Not reported per group. Average age: 43 years (range: 21-76). Duration sciatica: 7 days-7 weeks.	**Both groups received:** 3 visits. **I:** • Lumbar extension or lumbar extension with the hips offset. • 5 sets of 10 repetitions; 3 visits. **C:** • Mobilization.	**Pain** (Verbal analogue scale; 0-10 scale; the higher the worst): **I:** • Baseline: 3.9 ± 1.87. • After 3rd visit: 1.5 ± not reported. **C:** • Baseline: 3.8 ± 1.48. • After 3rd visit: 3.5 ± not reported. **Functional status** (ODI; 0-100 scale; the higher the worst). **I:** • Baseline: 27.40 ± 10.45. • After 3rd visit: 18.33 ± not reported. **C:** • Baseline: 29.50 ± 16.57. • After 3rd visit: 28.88 ± not reported. **Global perceived effect:** Not measured.	**Dropouts:** Not reported. **Side effects:** Not reported. **Adverse events:** Not reported.

Continued - Table 1. Study characteristics included studies

Thackeray et al. [34]	RCT. Measurement moments: • Baseline. • 8 weeks. • 6 months.	Herniated disc (MRI confirmed). Low back pain, with clinical and imaging findings consistent with lumbar disk herniation. Patients were scheduled to receive selective nerve root block. Total participants: N= 44 Gender: Not reported per group. Gender: 16 females (57%). Age: Not reported per group. Average age: 38.5 ± 11.6. Duration sciatica: 10 weeks (3-20 weeks)[a].	**Both groups received:** Injections (listed at control section). **I:** Exercise as an adjunct therapy to control: • Either a series of end-range directional exercises (N=15), or mechanical traction (N=6). • Strengthening, flexibility, stabilization, and cardio-vascular exercises, as directed by persistent impairments. • 4 weeks; average 6 sessions (range 2-13). **C:** Injections: • Maximum 3 injections. Between injections at least 2 weeks. • Selective Nerve Root Block (fluoroscopically guided; within 2 weeks of enrollment).	**Pain** (NRPS; 0 -10 scale; no pain -worst pain): **I:** • Baseline: 5.4 ± 2.5. • 8 weeks: 3.0 ± 2.3. • 6 months: 2.9 ± 2.5. **C:** • Baseline: 4.9 ± 2.0. • 8 weeks: 2.4 ± 2.4. • 6 months: 1.7 ± 2.5. **Functional status** (DSQI: 0-100 scale; no disability -incapacitated): **I:** • Baseline: 39.6 ± 21.6 • 8 weeks: 22.4 ± 18.3 • 6 months: 21.5 ± 19.6 **C:** • Baseline: 35.7 ± 16.7 • 8 weeks: 16.9 ± 18.2 • 6 months: 12.8 ± 19.5 **Global perceived effect** (GROC; -7 to +7 scale; a very great deal worse - a very great deal better): Effect: ≥4, 4 = 4= moderately better **I:** • 8 weeks: N = 13 (62%) • 6 months: N= 11 (52%) **C:** • 8 weeks: N= 15 (65%) • 6 months: N=18 (78%)	Pilot study Patients receiving physical therapy for low back pain, before joining the study: N=21 (48%). Small sample size. **Dropouts:** 4 (9%). **Side effects:** Not reported. **Adverse events:** Reported, but none found.

Outcome: $\bar{x}$ ± SD
[a] Outcome: median (interquartile range 25-75%)
I = Intervention group (exercise)
C = Control group

Risk of bias in included studies

Risk of bias was similar for all primary outcomes (Figure 2). Four studies performed an adequately blinded treatment allocation [18,23,19,34]. Blinding of participants, personnel, or outcome assessors, was impossible, due to the intervention and self-reported outcomes. For comparison to sham exercise, differences between groups were resolved by explaining benefits of both intervention and control. However, this explanation was not tested among patients, hence blinding of participants was regarded a high risk. Attrition bias was generally reported well and regarded as low risk. Reporting bias was regarded as low risk when reported [18,23,36,37,19]. Selection bias due to differences at baseline was regarded as high risk in the pilot study [34]. In most studies, risk of bias was poorly reported for the domains; co-interventions [18,23,36,32,37,33,35], compliance [18,23,36,32,37,33], and other forms of bias [23,36,32,37,33,35]. Co-interventions were reported in 2 RCTs: 1 regarded as high risk [34], and 1 as low risk [19]. RCTs reporting well on compliance were regarded as low risk [19,35,34]. The same applied for RCTs reporting on other forms of bias [18,19,34]. Intention to treat analysis was unclear in 3 RCTs [32,37,33]; others were regarded as low risk. Timing of outcome assessment was unclear in 3 RCTs [32,33,35] and low risk in 6 RCTs [18,23,36,37,19,34]. 1 study reported regarded as low risk, reported well on all items except blinding [19]. The risk of bias in the other studies was generally high.

Effects of interventions

7 RCTs used a simple parallel group design, 1 a factorial design [36] and 1 a crossover trial [23]. Pooled results and certainty of evidence are presented in table 2 and figure 3. One author was contacted for clarification since identical results were reported differently, [35]. Results were adjusted following correspondence with the author. Regarding functional status, 1 RCT reported the median instead of the mean [18]. Median was used as mean and SD was calculated.

Exercise compared to other therapies: 2 RCTs (317 participants) demonstrated a small, non-significant short-term effect on pain [SMD -0.48 (95% CI -1.45, 0.49); I^2 = 79%; very low

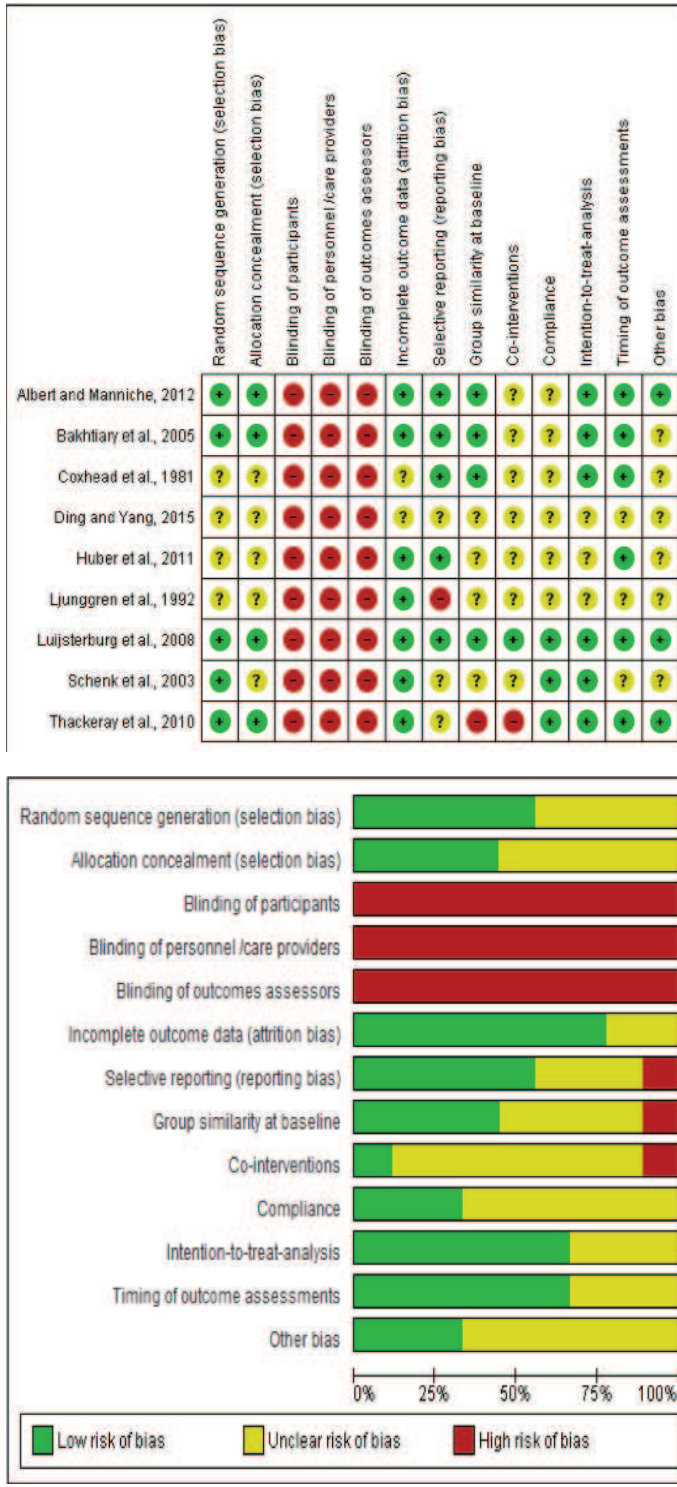


Figure 2. Risk of bias assessment of individual studies and across studies.

certainty; back-transformed MD -1.02 (95% CI -3.08, 1.04); 0-10 scale] [36,35]. The short-term effect of functional status was only measured in 1 small study that demonstrated a small, significant effect, [MD -8.4 (95% CI -15.70, -1.10); 0-100 scale; 25 participants; very low certainty] [35]. Short-term global perceived effect demonstrated a small, non-significant RR of 1.06 (95% CI 0.94, 1.19); 2 RCTs; 342 participants; I^2= 0%; very low certainty; RD 0.04; NNT=25) [36,33]. One study (250 participants) reported a small, non-significant intermediate-term effect on global perceived effect [RR 0.90 (95%CI 0.77, 1.05); very low certainty; RD -0.08;
NNT=13] [36]. None of the studies reported long-term effects.

Exercise compared to no therapy: The short-term effect of pain was large, significant and in favor of the intervention group [MD -2.07 (95% CI -3.24, -0.89); 0-10 scale; 2 RCTs; 112 participants; I^2 = 89%; very low certainty] [23,37]. No studies reported on short-term functional status, global perceived effect, and intermediate- or long-term effects.

Exercise compared to sham exercises: One study (181 participants) demonstrated a small, non-significant short-term effect on pain [MD -0.60 (95% CI -1.05, -0.15), 0-10 scale; very low certainty] and functional status [MD -0.50 (95% CI -1.48, 0.48); 0-24 scale; very low certainty] [18]. No short-term global perceived effects, nor intermediate-term effects were measured. Long-term effects showed a small, non-significant effect on pain [MD 0.30 (95% CI -0.14, 0.74), scale 0-10; very low certainty] and functional status [MD -0.50 (95% CI -1.68, 0.68); scale 0-24; very low certainty] [18]; global perceived effect was not measured.

Exercise as an adjunct therapy compared to the original therapy: The short-term effect of pain was medium and non-significant [SMD, -0.52 (95% CI -1.56, 0.52); 3 RCTs; 307 participants; I^2 = 94%; very low certainty; back-transformed MD -1.53 (95% CI -4.60, 1.53); 0-10 scale] [32,19,34]. The small short-term effect was non-significant for functional status [SMD -0.09 (95% CI -0.50, 0.31); 3 RCTs; 307 participants; I^2 = 66%; very low certainty; back-transformed MD -0.63 (95% CI -3.49, 2.17); 0-24 scale] and global perceived effect [RR 1.20 (95% CI 0.99,

Table 2. Summary of treatment effects and GRADE summary of primary analyses for all comparisons

Analyses	Effect estimate MD (95% CI) or RR (95% CI)	No of studies	No of partic-ipants
Exercise versus other therapies			
Pain (0-10 scale):			
Closest to 3 months	MD -1.02 (-3.08 to 1.04)*	2	317
Functional status scale (0-100 scale):			
Closest to 3 months	MD -8.4 (-15.70 to -1.10)	1	25
Global perceived effect:			
Closest to 3 months	RR 1.06 (0.94 to 1.19)	2	342
Closest to 6 months	RR 0.90 (0.77 to 1.05)	1	250
Exercise versus no therapy			
Pain (0-10 scale):			
Closest to 3 months	MD -2.07 (-3.24 to -0.89)	2	112
Exercise as an adjunct therapy versus the original therapy			
Pain (0-10 scale):			
Closest to 3 months	MD -1.53 (-4.60 to 1.53)*	3	307
Closest to 6 months	MD 0.70 (-0.22 to 1.62)	1	44
Closest to 12 months	MD -0.70 (-1.61 to 0.21)	1	135
Functional status scale:			
Closest to 3 months (0-24 scale)	MD -0.63 (-3.49 to 2.17)*	3	307
Closest to 6 months (0-100 scale)	MD 4.80 (-2.62 to 12.22)	1	44
Closest to 12 months (0-24 scale)	MD -0.90 (-3.03, 1.23)	1	135
Global perceived effect:			
Closest to 3 months	RR 1.20 (95% CI 0.99, 1.45)	3	307
Closest to 6 months	RR 0.67 (0.42 to 1.06)	1	44
Closest to 12 months	RR 1.42 (1.11 to 1.81)	1	135
Exercise versus sham exercises			
Pain (0-10 scale):			
Closest to 3 months	MD -0.60 (-1.05 to -0.15)	1	181
Closest to 12 months	MD 0.30 (-0.14 to 0.74)		
Functional status scale 0-24:			
Closest to 3 months	MD -0.50 (-1.48 to 0.48)	1	181
Closest to 12 months	MD -0.50 (-1.68 to 0.68)		

**MD back-transferred from SMD.*

*** Effect size was regarded as small, based on the original SMD.*

Continued - Table 2. Summary of treatment effects and GRADE summary of primary analyses for all comparisons

I^2 (%)	certainty of evidence (reason for downgrading)	Effect size
79	Very low certainty (limitations of design, imprecision, inconsistency, indirectness)	Small**
-	Very low certainty (limitations of design, imprecision, inconsistency)	Small
0	very low certainty (limitations of design, imprecision, indirectness)	Small
-	very low certainty (limitations of design, imprecision, inconsistency, indirectness)	Small
89	Very low certainty (limitations of design, imprecision, inconsistency)	Large
94	Very low certainty (limitations of design, imprecision, inconsistency)	Medium
-	Very low certainty (limitations of design, imprecision, inconsistency)	Small
-	Low certainty (imprecision, inconsistency)	Small
66	Very low certainty (limitations of design, imprecision, inconsistency)	Small
-	Very low certainty (limitations of design, imprecision, inconsistency)	Small
-	Low certainty (imprecision, inconsistency)	Small
23	Very low certainty (limitations of design, imprecision, inconsistency)	Small
-	Very low certainty (limitations of design, imprecision, inconsistency)	Medium
-	Low certainty (imprecision, inconsistency)	Medium
-	Very low certainty (imprecision, inconsistency, indirectness)	Small Small
-	Very low certainty (imprecision, inconsistency, indirectness)	Small Small

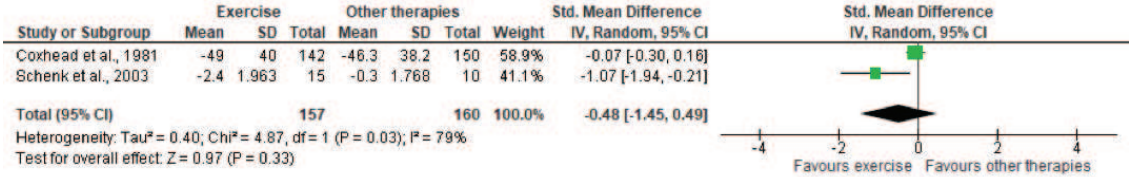
1.45); 3 RCTs; 307 participants; I^2 = 23%; low certainty; RD 0.12; NNT=8.33] [32,19,34]. One study (44 participants) reported on intermediate-term effects, where a small effect on pain [MD 0.70 (95% CI -0.22, 1.62); 0-10 scale; very low certainty], a small effect on functional status [MD 4.80 (95% CI -2.62, 12.22); 0-100 scale; very low certainty], and a medium effect on global perceived effect [RR 0.67 (95% CI 0.42, 1.06); very low certainty; RD -0.26; NNT=-4] were found; none were significant [34]. One study (135 participants) reported on the long-term effect; non-significant small effects were found on pain [MD -0.70 (95% CI -1.61, 0.21); 0-10 scale; low certainty], functional status [MD -0.90 (95% CI -3.03, 1.23); 0-24 scale; low certainty], and a significant medium effect was found on global perceived effect [RR 1.42 (95% CI 1.11, 1.81); RD 0.23; NNT=4; low certainty] [19]. Again SMD was calculated from SDs of Schenk et al. [35].

Secondary outcomes:
No significant short- and long-term effects were identified for neither quality of life [18,19] nor vocational outcomes [18,36,19]. Intermediate-effects were not measured. Physical outcomes were reported in four studies [18,23,32,37]. Different measurements were used (e.g. mobility, muscle strength, neurological signs) and contradicting results were found.

Subgroup analysis and investigation of heterogeneity
In order to explain heterogeneity subgroup analyses were performed for short-term effects. Subgroup analysis on intermediate and long term-effects were not possible due to the low number of studies. Subgroup analyses could not explain the heterogeneity.

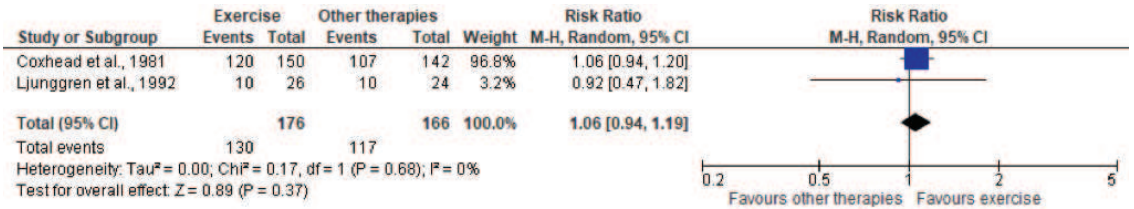
Sensitivity analysis
Sensitivity analyses did not change the effect size or the significance of the differences. However, for an informative sensitivity analysis more studies would be needed.

A. Exercise versus other therapies, pain, short-term.[a]



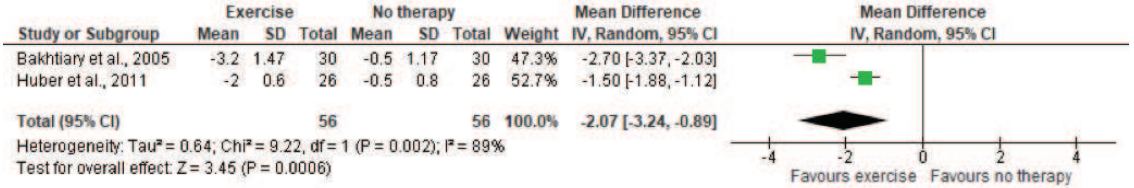
GRADE: very low certainty; downgraded for limitations of design, *imprecision, inconsistency, indirectness.*

B. Exercise versus other therapies, global perceived effect, short-term.[b]



GRADE: very low certainty; downgraded for limitations of design, imprecision, indirectness.imprecision, inconsistency, indirectness.

C. Exercise versus no therapy, pain, short-term.



GRADE: very low certainty; downgraded for limitations of design, imprecision, inconsistency.

Figure 3. Pooled short-, intermediate-, and long-term effects on pain differences and functional status (follow-up - baseline), differences in functional status (follow-up - baseline), and global perceived effect (number of patients perceiving effect). GRADE assessmewnt: high (⊕⊕⊕⊕), moderate (⊕⊕⊕⊙) low (⊕⊕⊙⊙), very low (⊕⊙⊙⊙).

Figure 3. Pooled short-, intermediate-, and long-term effects on pain differences and functional status (follow-up - baseline), differences in functional status (follow-up - baseline), and global perceived effect (number of patients perceiving effect). GRADE assessmewnt: high (⊕⊕⊕⊕), moderate (⊕⊕⊕⊙) low (⊕⊕⊙⊙), very low (⊕⊙⊙⊙).

D. Exercise as an adjunct therapy, pain, short-term. [a]

Study or Subgroup	Exercise Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
Ding and Yang, 2015	-7.9	2.4	64	-4.28	2.3	64	33.8%	-1.53 [-1.93, -1.14]
Luijsterburg et al., 2008	-3.9	2.8	67	-3.7	3.1	68	34.3%	-0.07 [-0.40, 0.27]
Thackeray et al., 2010	-2.4	1.5	21	-2.5	1.4	23	31.9%	0.07 [-0.52, 0.66]
Total (95% CI)			152			155	100.0%	-0.52 [-1.56, 0.52]

Heterogeneity: Tau² = 0.79; Chi² = 35.67, df = 2 (P < 0.00001); I² = 94%
Test for overall effect: Z = 0.98 (P = 0.33)

GRADE: very low certainty; downgraded for limitations of design, imprecision, inconsistency.

E. Exercise as an adjunct therapy, functional status, short-term. [a]

Study or Subgroup	Exercise Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
Ding and Yang, 2015	-3.7	2.3	64	-2.79	1.7	64	37.3%	-0.45 [-0.80, -0.10]
Luijsterburg et al., 2008	-7.7	7.3	67	-8.5	6.7	68	38.1%	0.11 [-0.22, 0.45]
Thackeray et al., 2010	-17.2	13	21	-18.8	11.1	23	24.6%	0.13 [-0.46, 0.72]
Total (95% CI)			152			155	100.0%	-0.09 [-0.50, 0.31]

Heterogeneity: Tau² = 0.08; Chi² = 5.86, df = 2 (P = 0.05); I² = 66%
Test for overall effect: Z = 0.44 (P = 0.66)

GRADE: very low certainty; downgraded for limitations of design, imprecision, inconsistency.

F. Exercise as an adjunct therapy, global perceived effect, short-term. [b]

Study or Subgroup	Exercise Events	Total	Control Events	Total	Weight	Risk Ratio M-H, Random, 95% CI
Ding and Yang, 2015	49	64	35	64	39.7%	1.40 [1.08, 1.82]
Luijsterburg et al., 2008	47	67	42	68	43.8%	1.14 [0.89, 1.45]
Thackeray et al., 2010	13	21	15	23	16.5%	0.95 [0.61, 1.49]
Total (95% CI)		152		155	100.0%	1.20 [0.99, 1.45]
Total events	109		92			

Heterogeneity: Tau² = 0.01; Chi² = 2.58, df = 2 (P = 0.28); I² = 23%
Test for overall effect: Z = 1.83 (P = 0.07)

GRADE: very low certainty; downgraded for limitations of design, imprecision, inconsistency.

[a] *Continuous outcomes were transformed to the same direction; a negative effect means a larger effect for exercise therapy. Transformation was performed in Coxhead et al. [36] and Ding, Yang [32] by multiplying the values by -1.[36,32]*

[b] *Values >1 means a larger effect for exercise therapy.*

DISCUSSION AND CONCLUSION

Summary of main results

Except for the comparison exercise compared to no therapy (large effect), only small effects were found. Due to the small amount of studies, sensitivity and subgroup analyses were difficult to perform and too few studies were available for long and mid-term effects analyses. The overall certainty of the evidence was low to very low and the high heterogeneity could not be explained.

Results compared to other literature

Our review is consistent with other findings. As in our review, Fernandez et al. [13] regarded the included studies as low-certainty studies. In contrast to our review, on the effect of pain, they found a small significant effect in favor of the exercise group compared to the advice group, whereas we found a large effect when compared to no therapy, and no significant effect compared to other therapies [13]. Although we have an overlap in RCTs, we excluded studies that compared one form of exercise to another. In addition, in a network meta-analyses of Lewis et al. [12], exercises had no significant effect on pain or global perceived effect when compared to other non-invasive therapies, but exercises were found to be inferior to epidural injections and intraoperative interventions for global perceived effect [12]. In this review, no results could be obtained for exercise compared to injections, but the findings of the effect of exercise compared to other therapies seems similar to our findings. They concluded that the lack of effect might be due to the small number of included studies [12]. To our knowledge there are no reviews reporting a large effect for exercises for sciatica.

Limitations

Most studies did not report on compliance. When compliance is low, intensity of exercise will decrease, and the estimated effect could be an underestimation. Furthermore, the results for pain should be viewed with caution. The distinction between leg and back pain was not possible because it was poorly reported. From some articles, it was unclear if pain referred to back or leg

pain. Sciatica is mainly characterized by radiating leg pain [7], therefore, if not reported we assumed that the pain patients referred to would be leg pain, but this could be debated.
We were not able to identify studies that examined the effect of exercise to general practitioner care, pharmacological treatments, injection or surgery, because this has not been explored.

Conclusion

Although, in clinical practice exercise therapy is often provided for sciatica, its effectiveness remains unclear. The overall certainty of the evidence was low to very low; sample sizes were small; and studies had a large unexplained statistical heterogeneity. More research in the form of larger, low risk of bias RCTs is highly recommended.

APPENDICES

APPENDIX I. SEARCH STRATEGY

Search strategy: key words

Types of studies
Controlled clinical trial, Randomized controlled trial, pragmatic clinical trial, randomized.

Types of participants
Sciatica, sciatic, lumbosacral radicular syndrome, radiculopathy, radiating leg pain, paresthesia leg pain, radicular pain, unilateral leg pain, low-back pain with neurological symptoms, dorsalgia, lumbago, ischias, ischialgia, lumboischialgia, radiculalgia, spinal/lumbar or sacral nerve root entrapment, spinal/lumbar or nerve roots entrapment, neuralgia, lumbar or sacral herniated disc, lumbar or sacral disc herniation, sciatic nerve entrapment, spondylosis, spondylitis, spondylolysis, slipped disc, slipped disc, slipped vertebra, Piriformis Syndrome, nerve compression syndromes, adults, humans.

Type of intervention
Exercise, exercises, exercise therapy, exercise movement techniques, physical training, physical exercises, physical activity, physical activities, aerobic exercises, anaerobic exercises, core stability, functional training, strength training, flexibility training, McKenzie, resistance training or exercises, graded activity, 3D training, sports, yoga, tai chi, endurance training, high intensity training, steps aerobics, swimming, aqua fitness, aquarobics, exercise movement techniques, physical fitness, Alexander, William, stretching exercises, mobility exercises, Mensendieck, Cesar, home exercises, hospital exercises, individual training, group training.

PubMed
#1 Sciatica filter
"Sciatic Neuropathy"[Mesh] OR sciatic*[tiab] OR Peroneal neuropath*[tiab] OR tibial neuropath*[tiab] OR ischias[tiab] OR

ischialg*[tiab] OR ischiasneuralg*[tiab] OR lumboischialg*[tiab]

pain and location:
((neuropath*[tiab] OR radiculopath*[tiab] OR radicular[tiab] OR radiculalg*[tiab] OR dorsalg*[tiab] OR neuralg*[tiab] OR spondylopath*[tiab]) AND (lumbosacral[tiab] OR lumbar[tiab] OR lumbal[tiab] OR sacral[tiab] OR leg[tiab] OR L4[tiab] OR L5[tiab] OR S1[tiab] OR S2[tiab]))

'Causes' of sciatica:
((“Intervertebral Disc Displacement”[Mesh] OR “Spinal Stenosis”[Mesh] OR “Spondylitis”[Mesh:NoExp] OR “Spondylosis”[Mesh] OR disc herniation[tiab] OR disk herniation[tiab] OR piriformis muscle syndrome*[tiab] OR piriformis syndrome*[tiab] OR pelvic tumor*[tiab] OR pelvic tumour*[tiab] OR nerve compression syndrome*[tiab] OR slipped disc*[tiab] OR slipped disk*[tiab] OR slipped vertebra*[tiab] OR intervertebral disk displacement[tiab] OR intervertebral disc displacement[tiab] OR herniated disc*[tiab] OR herniated disk*[tiab] OR prolapsed disc*[tiab] OR prolapsed disk*[tiab] OR nerve root entrapment[tiab] OR nerve roots entrapment[tiab] OR spinal stenosis[tiab] OR spondylosis[tiab] OR spondylitis[tiab] OR spondylolysis[tiab] OR spondylolisthesis[tiab]) AND (lumbosacral[tiab] OR sacral[tiab] OR lumbar[tiab] OR lumbal[tiab] OR L4[tiab] OR L5[tiab] OR S1[tiab] OR S2[tiab]))

<u>#2 Exercise filter</u>
“Motor Activity”[Mesh:NoExp] OR “Exercise”[Mesh] OR “Sports”[Mesh] OR “Physical Exertion”[Mesh] OR “Early Ambulation”[Mesh] OR “Exercise Therapy”[Mesh] OR “Exercise Movement Techniques”[Mesh] OR Motor Activit*[tiab] OR Physical Activit*[tiab] OR Locomotor Activit*[tiab] OR Exercis*[tiab] OR Physical Exercis*[tiab] OR Isometric Exercis*[tiab] OR Aerobic*[tiab] OR anaerobic*[tiab] OR core stability[tiab] OR training[tiab] OR stretching[tiab] OR Physical Condition*[tiab] OR Physical endurance[tiab] OR movement therap*[tiab] OR fitness[tiab] OR Plyometric[tiab] OR Stretch-Shortening[tiab] OR Weight-Lifting[tiab] OR Weight-Bearing[tiab] OR running[tiab] OR jogging[tiab] OR walk*[tiab]

OR bicycle[tiab] OR cycle[tiab] OR bicycling[tiab] OR cycling[tiab] OR rowing[tiab] OR swim*[tiab] OR ambulation[tiab] OR mobil*[tiab] OR pilates[tiab] OR yoga[tiab] OR McKenzie[tiab] OR alexander[tiab] OR William[tiab] OR graded activit*[tiab] OR tai chi[tiab] OR aquarobic*[tiab] OR mensendieck[tiab] OR cesar[tiab]

#3 RCT filter
(double-blind method[mh] OR single-blind method[mh] OR clinical trial[pt] OR "clinical trial"[tiab] OR "pragmatic trial"[tiab] OR "real world trial"[tiab] OR ((singl*[tiab] OR doubl*[tiab] OR trebl*[tiab] OR tripl*[tiab]) AND (mask*[tiab] OR blind*[tiab])) OR "latin square"[tiab] OR placebos[mh] OR placebo*[tiab] OR random*[tiab] OR research design[mh:noexp] OR comparative study[pt] OR evaluation studies[pt] OR follow-up studies[mh] OR prospective studies[mh] OR cross-over studies[mh] OR control[tiab] OR controll*[tiab] OR prospectiv*[tiab] OR volunteer*[tiab])

#4 humans filter
NOT (animals[mh] NOT humans[mh])

(#1 OR #2 OR #3) AND #4 AND #5

Search	PubMed Query 01-05-2019	Items found
#8	Search #7 NOT (animals[mh] NOT humans[mh])	**1725**
#7	Search #6 AND (double-blind method[mh] OR single-blind method[mh] OR clinical trial[pt] OR "clinical trial"[tiab] OR "pragmatic trial"[tiab] OR "real world trial"[tiab] OR ((singl*[tiab] OR doubl*[tiab] OR trebl*[tiab] OR tripl*[tiab]) AND (mask*[tiab] OR blind*[tiab])) OR "latin square"[tiab] OR placebos[mh] OR placebo*[tiab] OR random*[tiab] OR research design[mh:noexp] OR comparative study[pt] OR evaluation studies[pt] OR follow-up studies[mh] OR prospective studies[mh] OR cross-over studies[mh] OR control[tiab] OR controll*[tiab] OR prospectiv*[tiab] OR volunteer*[tiab])	2326

#6	Search #4 AND #5	5109
#5	Search "Motor Activity"[Mesh:NoExp] OR "Exercise"[Mesh] OR "Sports"[Mesh] OR "Physical Exertion"[Mesh] OR "Early Ambulation"[Mesh] OR "Exercise Therapy"[Mesh] OR "Exercise Movement Techniques"[Mesh] OR Motor Activit*[tiab] OR Physical Activit*[tiab] OR Locomotor Activit*[tiab] OR Exercis*[tiab] OR Physical Exercis*[tiab] OR Isometric Exercis*[tiab] OR Aerobic*[tiab] OR anaerobic*[tiab] OR core stability[tiab] OR training[tiab] OR stretching[tiab] OR Physical Condition*[tiab] OR Physical endurance[tiab] OR movement therap*[tiab] OR fitness[tiab] OR Plyometric[tiab] OR Stretch-Shortening[tiab] OR Weight-Lifting[tiab] OR Weight-Bearing[tiab] OR running[tiab] OR jogging[tiab] OR walk*[tiab] OR bicycle[tiab] OR cycle[tiab] OR bicycling[tiab] OR cycling[tiab] OR rowing[tiab] OR swim*[tiab] OR ambulation[tiab] OR mobil*[tiab] OR pilates[tiab] OR yoga[tiab] OR McKenzie[tiab] OR alexander[tiab] OR William[tiab] OR graded activit*[tiab] OR tai chi[tiab] OR aquarobic*[tiab] OR mensendieck[tiab] OR cesar[tiab]	1914893
#4	Search #1 OR #2 OR #3	55234
#3	Search (("Intervertebral Disc Displacement"[Mesh] OR "Spinal Stenosis"[Mesh] OR "Spondylitis"[Mesh:NoExp] OR "Spondylosis"[Mesh] OR disc herniation[tiab] OR disk herniation[tiab] OR piriformis muscle syndrome*[tiab] OR piriformis syndrome*[tiab] OR pelvic tumor*[tiab] OR pelvic tumour*[tiab] OR nerve compression syndrome*[tiab] OR slipped disc*[tiab] OR slipped disk*[tiab] OR slipped vertebra*[tiab] OR intervertebral disk displacement[tiab] OR intervertebral disc displacement[tiab] OR herniated disc*[tiab] OR herniated disk*[tiab] OR prolapsed disc*[tiab] OR prolapsed disk*[tiab] OR nerve root entrapment[tiab] OR nerve roots entrapment[tiab] OR spinal stenosis[tiab] OR spondylosis[tiab] OR spondylitis[tiab] OR spondylolysis[tiab] OR spondylolisthesis[tiab]) AND (lumbosacral[tiab] OR sacral[tiab] OR lumbar[tiab] OR lumbal[tiab] OR L4[tiab] OR L5[tiab] OR S1[tiab] OR S2[tiab]))	20280
#2	Search ((neuropath*[tiab] OR radiculopath*[tiab] OR radicular[tiab] OR radiculalg*[tiab] OR dorsalg*[tiab] OR neuralg*[tiab] OR spondylopath*[tiab]) AND (lumbosacral[tiab] OR lumbar[tiab] OR lumbal[tiab] OR sacral[tiab] OR leg[tiab] OR L4[tiab] OR L5[tiab] OR S1[tiab] OR S2[tiab]))	10106
#1	Search "Sciatic Neuropathy"[Mesh] OR sciatic*[tiab] OR Peroneal neuropath*[tiab] OR tibial neuropath*[tiab] OR ischias[tiab] OR ischialg*[tiab] OR ischiasneuralg*[tiab] OR lumboischialg*[tiab]	30272

EMBASE.com

#1 Sciatica filter

'sciatic neuropathy'/exp OR sciatic*:ti,ab OR 'Peroneal neuropath*':ti,ab OR 'tibial neuropath*':ti,ab OR ischias:ti,ab OR ischialg*:ti,ab OR ischiasneuralg*:ti,ab OR lumboischialg*:ti,ab

pain and location:

((neuropath* OR radiculopath* OR radicular OR radiculalg* OR dorsalg* OR neuralg* OR spondylopath*) NEAR/3 (lumbosacral OR sacral OR lumbal OR lumbar OR leg OR l4 OR l5 OR s1 OR s2)):ti,ab

'Causes' of sciatica:

(('intervertebral disk hernia'/exp OR 'vertebral canal stenosis'/exp OR 'spondylitis'/de OR 'spondylosis'/de OR 'disc herniation:ti,ab' OR 'disk herniation':ti,ab OR 'piriformis muscle syndrome*':ti,ab OR 'piriformis syndrome*':ti,ab OR 'pelvic tumor*':ti,ab OR 'pelvic tumour*':ti,ab OR 'nerve compression syndrome*':ti,ab OR 'slipped disc*':ti,ab OR 'slipped disk*':ti,ab OR 'slipped vertebra*':ti,ab OR 'intervertebral disk displacement':ti,ab OR 'intervertebral disc displacement':ti,ab OR 'herniated disc*':ti,ab OR 'herniated disk*':ti,ab OR 'prolapsed disc*':ti,ab OR 'prolapsed disk*':ti,ab OR 'nerve root entrapment':ti,ab OR 'nerve roots entrapment':ti,ab OR 'spinal stenosis':ti,ab OR 'spondylosis':ti,ab OR 'spondylitis':ti,ab OR 'spondylolysis':ti,ab OR spondylolisthesis:ti,ab) AND ('lumbosacral':ti,ab OR 'lumbar':ti,ab OR lumbal:ti,ab OR sacral:ti,ab OR L4:ti,ab OR L5:ti,ab OR S1:ti,ab OR S2:ti,ab))

#2 Exercise filter

'motor activity'/de OR 'exercise'/exp OR 'sport'/exp OR 'mobilization'/exp OR 'kinesiotherapy'/exp OR 'Motor Activit*':ti,ab OR 'Physical Activit*':ti,ab OR 'Locomotor Activit*':ti,ab OR Exercis*:ti,ab OR 'Physical Exercis*':ti,ab OR 'Isometric Exercis*':ti,ab OR Aerobic*:ti,ab OR anaerobic*:ti,ab OR 'core stability':ti,ab OR training:ti,ab OR stretching:ti,ab OR 'Physical Condition*':ti,ab OR 'Physical endurance':ti,ab OR 'movement therap*':ti,ab OR fitness:ti,ab OR Plyometric:ti,ab OR 'Stretch-Shortening':ti,ab OR 'Weight-Lifting':ti,ab OR 'Weight-Bearing':ti,ab OR running:ti,ab OR jogging:ti,ab OR

walk*:ti,ab OR bicycle:ti,ab OR cycle:ti,ab OR bicycling:ti,ab OR cycling:ti,ab OR rowing:ti,ab OR swim*:ti,ab OR ambulation:ti,ab OR mobil*:ti,ab OR pilates:ti,ab OR yoga:ti,ab OR McKenzie:ti,ab OR alexander:ti,ab OR William:ti,ab OR 'graded activit*':ti,ab OR 'tai chi':ti,ab OR aquarobic*:ti,ab OR mensendieck:ti,ab OR cesar:ti,ab

#3 RCT filter
('article'/it OR 'article in press'/it OR 'review'/it) AND ('clinical trial'/de OR 'comparative study'/de OR 'controlled clinical trial'/de OR 'controlled study'/de OR 'human'/de OR 'intermethod comparison'/de OR 'methodology'/de OR 'prospective study'/de OR 'randomized controlled trial'/de OR 'systematic review'/de)

#4 humans filter
NOT ([animals]/lim NOT [humans]/lim)

Embase Session Results (01-05-2019)

No.	Query	Results
#8	#7 NOT ([animals]/lim NOT [humans]/lim)	**3089**
#7	#6 AND ('article'/it OR 'article in press'/it OR 'review'/it) AND ('clinical trial'/de OR 'comparative study'/de OR 'controlled clinical trial'/de OR 'controlled study'/de OR 'human'/de OR 'intermethod comparison'/de OR 'methodology'/de OR 'prospective study'/de OR 'randomized controlled trial'/de OR 'systematic review'/de)	4041
#6	#4 AND #5	6389
#5	'motor activity'/de OR 'exercise'/exp OR 'sport'/exp OR 'mobilization'/exp OR 'kinesiotherapy'/exp OR 'motor activit*':ti,ab OR 'physical activit*':ti,ab OR 'locomotor activit*':ti,ab OR exercis*:ti,ab OR 'physical exercis*':ti,ab OR 'isometric exercis*':ti,ab OR aerobic*:ti,ab OR anaerobic*:ti,ab OR 'core stability':ti,ab OR training:ti,ab OR stretching:ti,ab OR 'physical condition*':ti,ab OR 'physical endurance':ti,ab OR 'movement therap*':ti,ab OR fitness:ti,ab OR plyometric:ti,ab OR 'stretch-shortening':ti,ab OR 'weight-lifting':ti,ab OR 'weight-bearing':ti,ab OR running:ti,ab OR jogging:ti,ab OR walk*:ti,ab OR bicycle:ti,ab OR cycle:ti,ab OR bicycling:ti,ab OR cycling:ti,ab OR rowing:ti,ab OR swim*:ti,ab OR ambulation:ti,ab OR mobil*:ti,ab OR pilates:ti,ab OR yoga:ti,ab OR mckenzie:ti,ab OR alexander:ti,ab OR william:ti,ab OR 'graded activit*':ti,ab OR 'tai chi':ti,ab OR aquarobic*:ti,ab OR mensendieck:ti,ab OR cesar:ti,ab	2397932
#4	#1 OR #2 OR #3	60941

#3	('intervertebral disk hernia'/exp OR 'vertebral canal stenosis'/exp OR 'spondylitis'/de OR 'spondylosis'/de OR 'disc herniation:ti,ab' OR 'disk herniation':ti,ab OR 'piriformis muscle syndrome*':ti,ab OR 'piriformis syndrome*':ti,ab OR 'pelvic tumor*':ti,ab OR 'pelvic tumour*':ti,ab OR 'nerve compression syndrome*':ti,ab OR 'slipped disc*':ti,ab OR 'slipped disk*':ti,ab OR 'slipped vertebra*':ti,ab OR 'intervertebral disk displacement':ti,ab OR 'intervertebral disc displacement':ti,ab OR 'herniated disc*':ti,ab OR 'herniated disk*':ti,ab OR 'prolapsed disc*':ti,ab OR 'prolapsed disk*':ti,ab OR 'nerve root entrapment':ti,ab OR 'nerve roots entrapment':ti,ab OR 'spinal stenosis':ti,ab OR 'spondylosis':ti,ab OR 'spondylitis':ti,ab OR 'spondylolysis':ti,ab OR spondylolisthesis:ti,ab) AND ('lumbosacral':ti,ab OR 'lumbar':ti,ab OR lumbal:ti,ab OR sacral:ti,ab OR l4:ti,ab OR l5:ti,ab OR s1:ti,ab OR s2:ti,ab)	25603
#2	((neuropath* OR radiculopath* OR radicular OR radiculalg* OR dorsalg* OR neuralg* OR spondylopath*) NEAR/3 (lumbosacral OR sacral OR lumbal OR lumbar OR leg OR l4 OR l5 OR s1 OR s2)):ti,ab	3705
#1	'sciatic neuropathy'/exp OR sciatic*:ti,ab OR 'peroneal neuropath*':ti,ab OR 'tibial neuropath*':ti,ab OR ischias:ti,ab OR ischialg*:ti,ab OR ischiasneuralg*:ti,ab OR lumboischialg*:ti,ab	34782

PEDro database
#1 Sciatica filter
Sciatic*

#2 RCT filter
Trials

PEDro results (01-05-2019)
Results: 94

Cinahl
#1 Sciatica filter
(MH "Sciatic Nerve") OR (MH "Sciatica") OR TI (sciatic* OR Peroneal neuropath* OR tibial neuropath* OR ischias OR ischialg* OR ischiasneuralg* OR lumboischialg*) OR AB (sciatic* OR Peroneal neuropath* OR tibial neuropath* OR ischias OR ischialg* OR ischiasneuralg* OR lumboischialg*)

pain and location:
(TI (neuropath* OR radiculopath* OR radicular OR radiculalg* OR dorsalg* OR neuralg* OR spondylopath*) AND TI (lumbosacral

OR lumbar OR lumbal OR sacral OR leg OR L4 OR L5 OR S1 OR S2)) OR (AB (neuropath* OR radiculopath* OR radicular OR radiculalg* OR dorsalg* OR neuralg* OR spondylopath*) AND AB (lumbosacral OR lumbar OR lumbal OR sacral OR leg OR L4 OR L5 OR S1 OR S2))

'Causes' of sciatica:
(((MH "Intervertebral Disk Displacement") OR (MH "Spinal Stenosis") OR (MH "Spondylitis, Ankylosing") OR (MH "Spondylosis") OR (MH "Spondylolysis") OR (MH "Spondylolisthesis") OR TI (disc herniation OR disk herniation OR piriformis muscle syndrome* OR piriformis syndrome* OR pelvic tumor* OR pelvic tumour* OR nerve compression syndrome* OR slipped disc* OR slipped disk* OR slipped vertebra* OR intervertebral disk displacement OR intervertebral disc displacement OR herniated disc* OR herniated disk* OR prolapsed disc* OR prolapsed disk*OR nerve root entrapment OR nerve roots entrapment OR spinal stenosis OR spondylosis OR spondylitis OR spondylolysis OR spondylolisthesis) OR AB (disc herniation OR disk herniation OR piriformis muscle syndrome* OR piriformis syndrome* OR pelvic tumor* OR pelvic tumour* OR nerve compression syndrome* OR slipped disc* OR slipped disk* OR slipped vertebra* OR intervertebral disk displacement OR intervertebral disc displacement OR herniated disc* OR herniated disk* OR prolapsed disc* OR prolapsed disk*OR nerve root entrapment OR nerve roots entrapment OR spinal stenosis OR spondylosis OR spondylitis OR spondylolysis OR spondylolisthesis)) AND (TI (lumbosacral OR sacral OR lumbar OR lumbal OR L4 OR L5 OR S1 OR S2) OR AB (lumbosacral OR sacral OR lumbar OR lumbal OR L4 OR L5 OR S1 OR S2))

#2 Exercise filter
(MH "Motor Activity") OR (MH "Exercise") OR (MH "Sports+") OR (MH "Exertion") OR (MH "Early Ambulation") OR (MH "Therapeutic Exercise+") OR TI (Motor Activit* OR Physical Activit* OR Locomotor Activit* OR Exercis* OR Physical Exercis* OR Isometric Exercis* OR Aerobic* OR anaerobic* OR core stability OR training OR stretching OR Physical Condition* OR Physical endurance OR movement therap* OR fitness OR Plyometric OR Stretch-Shortening OR Weight-Lifting OR Weight-

Bearing OR running OR jogging OR walk* OR bicycle OR cycle OR bicycling OR cycling OR rowing OR swim* OR ambulation OR mobil* OR pilates OR yoga OR McKenzie OR alexander OR William OR graded activit* OR tai chi OR aquarobic* OR mensendieck OR cesar) OR AB (Motor Activit* OR Physical Activit* OR Locomotor Activit* OR Exercis* OR Physical Exercis* OR Isometric Exercis* OR Aerobic* OR anaerobic* OR core stability OR training OR stretching OR Physical Condition* OR Physical endurance OR movement therap* OR fitness OR Plyometric OR Stretch-Shortening OR Weight-Lifting OR Weight-Bearing OR running OR jogging OR walk* OR bicycle OR cycle OR bicycling OR cycling OR rowing OR swim* OR ambulation OR mobil* OR pilates OR yoga OR McKenzie OR alexander OR William OR graded activit* OR tai chi OR aquarobic* OR mensendieck OR cesar)

#3 RCT filter filter
(MH "Clinical Trials+") OR (PT Clinical trial) OR (TX clini* N1 trial*) OR (TX ((singl* N1 blind*) or (singl* N1 mask*)) or TX ((doubl* N1 blind*) or (doubl* N1 mask*)) OR or TX ((tripl* N1 blind*) or (tripl* N1 mask*))) OR (TX randomi* control*) OR (MH "Random Assignment") OR ((TX random* allocat*) or (TX allocat* random*)) OR (TX placebo*) OR (TX (waitlist* or (wait* and list*)) and (control* or group))) OR ((TX "treatment as usual") or (TX tau)) OR (TX (control* N3 (trial* or study or studies or group*))) OR (MH "Quantitative Studies")

#4 humans filter
NOT (MH "Animals" NOT MH "Human)

#	Cinahl Query 01-05-2019	Results
S9	S8 NOT (MH "Animals" NOT MH "Human)	**379**
S8	S6 AND S7	383
S7	(MH "Clinical Trials+") OR (PT Clinical trial) OR (TX clini* N1 trial*) OR (TX ((singl* N1 blind*) or (singl* N1 mask*)) or TX ((doubl* N1 blind*) or (doubl* N1 mask*)) OR or TX ((tripl* N1 blind*) or (tripl* N1 mask*))) OR (TX randomi* control*) OR (MH "Random Assignment") OR ((TX random* allocat*) or (TX allocat* random*)) OR (TX placebo*) OR (TX (waitlist* or (wait* and list*)) and (control* or group))) OR ((TX "treatment as usual") or (TX tau)) OR (TX (control* N3 (trial* or study or studies or group*))) OR (MH "Quantitative Studies")	575,457
S6	S4 AND S5	1,644

S5	(MH “Motor Activity”) OR (MH “Exercise”) OR (MH “Sports+”) OR (MH “Exertion”) OR (MH “Early Ambulation”) OR (MH “Therapeutic Exercise+”) OR TI (Motor Activit* OR Physical Activit* OR Locomotor Activit* OR Exercis* OR Physical Exercis* OR Isometric Exercis* OR Aerobic* OR anaerobic* OR core stability OR training OR stretching OR Physical Condition* OR Physical endurance OR movement therap* OR fitness OR Plyometric OR Stretch-Shortening OR Weight-Lifting OR Weight-Bearing OR running OR jogging OR walk* OR bicycle OR cycle OR bicycling OR cycling OR rowing OR swim* OR ambulation OR mobil* OR pilates OR yoga OR McKenzie OR alexander OR William OR graded activit* OR tai chi OR aquarobic* OR mensendieck OR cesar) OR AB (Motor Activit* OR Physical Activit* OR Locomotor Activit* OR Exercis* OR Physical Exercis* OR Isometric Exercis* OR Aerobic* OR anaerobic* OR core stability OR training OR stretching OR Physical Condition* OR Physical endurance OR movement therap* OR fitness OR Plyometric OR Stretch-Shortening OR Weight-Lifting OR Weight-Bearing OR running OR jogging OR walk* OR bicycle OR cycle OR bicycling OR cycling OR rowing OR swim* OR ambulation OR mobil* OR pilates OR yoga OR McKenzie OR alexander OR William OR graded activit* OR tai chi OR aquarobic* OR mensendieck OR cesar)	500,530
S4	S1 OR S2 OR S3	11,083
S3	(((MH “Intervertebral Disk Displacement”) OR (MH “Spinal Stenosis”) OR (MH “Spondylitis, Ankylosing”) OR (MH “Spondylosis”) OR (MH “Spondylolysis”) OR (MH “Spondylolisthesis”) OR TI (disc herniation OR disk herniation OR piriformis muscle syndrome* OR piriformis syndrome* OR pelvic tumor* OR pelvic tumour* OR nerve compression syndrome* OR slipped disc* OR slipped disk* OR slipped vertebra* OR intervertebral disk displacement OR intervertebral disc displacement OR herniated disc* OR herniated disk* OR prolapsed disc* OR prolapsed disk*OR nerve root entrapment OR nerve roots entrapment OR spinal stenosis OR spondylosis OR spondylitis OR spondylolysis OR spondylolisthesis) OR AB (disc herniation OR disk herniation OR piriformis muscle syndrome* OR piriformis syndrome* OR pelvic tumor* OR pelvic tumour* OR nerve compression syndrome* OR slipped disc* OR slipped disk* OR slipped vertebra* OR intervertebral disk displacement OR intervertebral disc displacement OR herniated disc* OR herniated disk* OR prolapsed disc* OR prolapsed disk*OR nerve root entrapment OR nerve roots entrapment OR spinal stenosis OR spondylosis OR spondylitis OR spondylolysis OR spondylolisthesis)) AND (TI (lumbosacral OR sacral OR lumbar OR lumbal OR L4 OR L5 OR S1 OR S2) OR AB (lumbosacral OR sacral OR lumbar OR lumbal OR L4 OR L5 OR S1 OR S2))	5,640
S2	(TI (neuropath* OR radiculopath* OR radicular OR radiculalg* OR dorsalg* OR neuralg* OR spondylopath*) AND TI (lumbosacral OR lumbar OR lumbal OR sacral OR leg OR L4 OR L5 OR S1 OR S2)) OR (AB (neuropath* OR radiculopath* OR radicular OR radiculalg* OR dorsalg* OR neuralg* OR spondylopath*) AND AB (lumbosacral OR lumbar OR lumbal OR sacral OR leg OR L4 OR L5 OR S1 OR S2))	3,194

S1	(MH "Sciatic Nerve") OR (MH "Sciatica") OR TI (sciatic* OR Peroneal neuropath* OR tibial neuropath* OR ischias OR ischialg* OR ischiasneuralg* OR lumboischialg*) OR AB (sciatic* OR Peroneal neuropath* OR tibial neuropath* OR ischias OR ischialg* OR ischiasneuralg* OR lumboischialg*)	4,505

Cochrane Library

#1 Sciatica

sciatic* OR "Peroneal neuropath*" OR "tibial neuropath*" OR ischias OR ischialg* OR ischiasneuralg* OR lumboischialg*

pain and location:

((neuropath* OR radiculopath* OR radicular OR radiculalg* OR dorsalg* OR neuralg* OR spondylopath*) AND (lumbosacral OR lumbar OR lumbal OR leg OR sacral OR L4 OR L5 OR S1 OR S2))

'Causes' of sciatica:

(("disc herniation" OR "disk herniation" OR "piriformis muscle syndrome*" OR "piriformis syndrome*" OR "pelvic tumor*" OR "pelvic tumour*" OR "nerve compression syndrome*" OR "slipped disc*" OR "slipped disk*" OR "slipped vertebra*" OR "intervertebral disk displacement" OR "intervertebral disc displacement" OR "herniated disc*" OR "herniated disk*" OR "prolapsed disc*" OR "prolapsed disk*" OR "nerve root entrapment" OR "nerve roots entrapment" OR "spinal stenosis" OR spondylosis OR spondylitis OR spondylolysis OR spondylolisthesis) AND (lumbosacral OR lumbar OR lumbal OR sacral OR L4 OR L5 OR S1 OR S2))

#2 Exercise filter

Sports OR sport OR "Physical Exertion" OR "Early Ambulation" OR "Exercise Therapy" OR "Exercise Movement Techniques" OR "Motor Activit*" OR "Physical Activit*" OR "Locomotor Activit*" OR Exercis* OR "Physical Exercis*" OR "Isometric Exercis*" OR Aerobic* OR anaerobic* OR "core stability" OR training OR stretching OR "Physical Condition*" OR "Physical endurance" OR "movement therap*" OR fitness OR Plyometric OR "Stretch-Shortening" OR "Weight-Lifting" OR "Weight-Bearing" OR running OR jogging OR walk* OR bicycle OR cycle OR bicycling OR cycling OR rowing OR swim* OR ambulation OR mobil* OR pilates OR yoga OR McKenzie OR alexander OR William OR "graded activit*" OR tai chi OR aquarobic* OR mensendieck OR cesar

#1	(sciatic* OR "Peroneal neuropath*" OR "tibial neuropath*" OR ischias OR ischialg* OR ischiasneuralg* OR lumboischialg*):ti,ab,kw	S	Limits	1868
#2	(((neuropath* OR radiculopath* OR radicular OR radiculalg* OR dorsalg* OR neuralg* OR spondylopath*) AND (lumbosacral OR lumbar OR lumbal OR leg OR sacral OR L4 OR L5 OR S1 OR S2))):ti,ab,kw	S	Limits	1593
#3	(("disc herniation" OR "disk herniation" OR "piriformis muscle syndrome*" OR "piriformis syndrome*" OR "pelvic tumor*" OR "pelvic tumour*" OR "nerve compression syndrome*" OR "slipped disc*" OR "slipped disk*" OR "slipped vertebra*" OR "intervertebral disk displacement" OR "intervertebral disc displacement" OR "herniated disc*" OR "herniated disk*" OR "prolapsed disc*" OR "prolapsed disk*" OR "nerve root entrapment" OR "nerve roots entrapment" OR "spinal stenosis" OR spondylosis OR spondylitis OR spondylolysis OR spondylolisthesis) AND (lumbosacral OR lumbar OR lumbal OR sacral OR L4 OR	S	Limits	2573
#4	(Sports OR sport OR "Physical Exertion" OR "Early Ambulation" OR "Exercise Therapy" OR "Exercise Movement Techniques" OR "Motor Activit*" OR "Physical Activit*" OR "Locomotor Activit*" OR Exercis* OR "Physical Exercis*" OR "Isometric Exercis*" OR Aerobic* OR anaerobic* OR "core stability" OR training OR stretching OR "Physical Condition*" OR "Physical endurance" OR "movement therap*" OR fitness OR Plyometric OR "Stretch-Shortening" OR "Weight-Lifting" OR "Weight-Bearing" OR running OR jogging OR walk* OR bicycle OR cycle OR bicycling OR cycling OR rowing OR swim* OR ambulation OR mobil* OR pilates OR yoga OR McKenzie OR alexander OR William OR "graded activit*" OR tai chi OR aquarobic* OR mensendieck OR cesar):ti,ab,kw	S	Limits	198477
#5	#1 OR #2 OR #3		Limits	5257
#6	#4 AND #5		Limits	1100

in Cochrane Reviews, Cochrane Protocols and Trials

APPENDIX II. DOMAINS OF RISK OF BIAS AND CRITERIA FOR JUDGEMENT FOR THE SOURCES OF RISK OF BIAS.

Table 3 Domains and source of Bias according the paper of Furlan et al. [22].

Bias Domain	Source of Bias	Possible Answers
Selection	(1) Was the method of randomization adequate?	Yes/No/Unsure
Selection	(2) Was the treatment allocation concealed?	Yes/No/Unsure
Performance	(3) Was the patient blinded to the intervention?	Yes/No/Unsure
Performance	(4) Was the care provider blinded to the intervention?	Yes/No/Unsure
Detection	(5) Was the outcome assessor blinded to the intervention?	Yes/No/Unsure
Attrition	(6) Was the drop-out rate described and acceptable?	Yes/No/Unsure
Attrition	(7) Were all randomized participants analyzed in the group to which they were allocated?	Yes/No/Unsure
Reporting	(8) Are reports of the study free of suggestion of selective outcome reporting?	Yes/No/Unsure
Selection	(9) Were the groups similar at baseline regarding the most important prognostic indicators?	Yes/No/Unsure
Performance	(10) Were cointerventions avoided or similar?	Yes/No/Unsure
Performance	(11) Was the compliance acceptable in all groups?	Yes/No/Unsure
Detection	(12) Was the timing of the outcome assessment similar in all groups?	Yes/No/Unsure
Other	(13) Are other sources of potential bias unlikely?	Yes/No/Unsure

Table 4.'Criteria for a Judgment of "Yes" for the Sources of Risk of Bias' according to the paper of Furlan et al. [22].

1	A random (unpredictable) assignment sequence. Examples of adequate methods are coin toss (for studies with 2 groups), rolling a dice (for studies with 2 or more groups), drawing of balls of different colors, drawing of ballots with the study group labels from a dark bag, computer-generated random sequence, preordered sealed envelopes, sequentially-ordered vials, telephone call to a central office, and preordered list of treatment assignments. Examples of inadequate methods are: alternation, birth date, social insurance/security number, date in which they are invited to participate in the study, and hospital registration number.
2	Assignment generated by an independent person not responsible for determining the eligibility of the patients. This person has no information about the persons included in the trial and has no influence on the assignment sequence or on the decision about eligibility of the patient.
3	Index and control groups are indistinguishable for the patients or if the success of blinding was tested among the patients and it was successful.

4	Index and control groups are indistinguishable for the care providers or if the success of blinding was tested among the care providers and it was successful.
5	Adequacy of blinding should be assessed for each primary outcome separately. This item should be scored "yes" if the success of blinding was tested among the outcome assessors and it was successful or:
	• for patient-reported outcomes in which the patient is the outcome assessor (e.g., pain, disability): the blinding procedure is adequate for outcome assessors if participant blinding is scored "yes"
	• for outcome criteria assessed during scheduled visit and that supposes a contact between participants and outcome assessors (e.g., clinical examination): the blinding procedure is adequate if patients are blinded, and the treatment or adverse effects of the treatment cannot be noticed during clinical examination
	• for outcome criteria that do not suppose a contact with participants (e.g., radiography, magnetic resonance imaging): the blinding procedure is adequate if the treatment or adverse effects of the treatment cannot be noticed when assessing the main outcome
	• for outcome criteria that are clinical or therapeutic events that will be determined by the interaction between patients and care providers (e.g., cointerventions, hospitalization length, treatment failure), in which the care provider is the outcome assessor: the blinding procedure is adequate for outcome assessors if item "4" (caregivers) is scored "yes"
	• for outcome criteria that are assessed from data of the medical forms: the blinding procedure is adequate if the treatment or adverse effects of the treatment cannot be noticed on the extracted data
6	The number of participants who were included in the study but did not complete the observation period or were not included in the analysis must be described and reasons given. If the percentage of withdrawals and drop-outs does not exceed 20% for short-term follow-up and 30% for long-term follow-up and does not lead to substantial bias a "yes" is scored. (N.B. these percentages are arbitrary, not supported by literature).
7	All randomized patients are reported/analyzed in the group they were allocated to by randomization for the most important moments of effect measurement (minus missing values) irrespective of noncompliance and cointerventions.
8	All the results from all prespecified outcomes have been adequately reported in the published report of the trial. This information is either obtained by comparing the protocol and the report, or in the absence of the protocol assessing that the published report includes enough information to make this judgment.
9	Groups have to be similar at baseline regarding demographic factors, duration and severity of complaints, percentage of patients with neurological symptoms, and value of main outcome measure(s).
10	If there were no cointerventions or they were similar between the index and control groups.
11	The reviewer determines if the compliance with the interventions is acceptable, based on the reported intensity, duration, number and frequency of sessions for both the index intervention and control intervention(s). For example, physiotherapy treatment is usually administered for several sessions; therefore it is necessary to assess how many sessions each patient attended. For single-session interventions (e.g., surgery), this item is irrelevant.

12	Timing of outcome assessment should be identical for all intervention groups and for all primary outcome measures.
13	Other types of biases. For example:
	• When the outcome measures were not valid. There should be evidence from a previous or present scientific study that the primary outcome can be considered valid in the context of the present. • Industry-sponsored trials. The conflict of interest (COI) statement should explicitly state that the researchers have had full possession of the trial process from planning to reporting without funders with potential COI having any possibility to interfere in the process. If, for example, the statistical analyses have been done by a funder with a potential COI, usually "unsure" is scored.

APPENDIX III. HETEROGENEITY CLASSIFICATION.

Statistical heterogeneity was classified according to Cochrane handbook [26].

- 0% to 40%: might not be important;
- 30% to 60%: may represent moderate heterogeneity*;
- 50% to 90%: may represent substantial heterogeneity*;
- 75% to 100%: considerable heterogeneity*.

*The importance of the observed value of I2 depends on (i) magnitude and direction of effects and (ii) strength of evidence for heterogeneity (e.g. P value from the chi-squared test, or a confidence interval for I2).

APPENDIX IV. GRADE DEFINITIONS.

The GRADE definitions of the Cochrane Back and Neck group [22,27].

- **High** (⊕⊕⊕⊕): Further research is very unlikely to change the confidence in the estimate of effect.
- **Moderate** (⊕⊕⊕⊙): Further research is likely to have an important impact in the confidence in the estimate of effect and may change the estimate.
- **Low** (⊕⊕⊙⊙): Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
- **Very low** (⊕⊙⊙⊙): Any estimate of effect is very uncertain

REFERENCES

1. Heliovaara M, Impivaara O, Sievers K, Melkas T, Knekt P, Korpi J, Aromaa A (1987) Lumbar disc syndrome in Finland. *J Epidemiol Community Health 41* (3):251-258.

2. Konstantinou K, Hider SL, Jordan JL, Lewis M, Dunn KM, Hay EM (2013) The impact of low back-related leg pain on outcomes as compared with low back pain alone: a systematic review of the literature. *Clin J Pain 29* (7):644-654.

3. Oosterhuis T, Smaardijk V, Kuijer P, Hoving J, Frings-Dresen M (2016) Actualisatie wetenschappelijke kennis en inzichten over het lumbosacraal radiculair syndroom. [Internet]. Available from: https://www.uwv.nl/overuwv/Images/Actualisatie-lumbosacraal-radiculair-syndroom.pdf. Accessed 2019 April 16th. Dutch.

4. Schaafstra A, Spinnewijn WEM, Bons S, Borg M, Koes BW, Ostelo R, Spijker-Huiges A, Burgers J, Bouma M, Verburg AFE (2015) NHG-Standaard Lumbosacraal radiculair syndroom (tweede herziening). *Huisarts & Wetenschap 58* (6):308-320. Dutch.

5. Spijker-Huiges A, Groenhof F, Winters JC, van Wijhe M, Groenier KH, van der Meer K (2015) Radiating low back pain in general practice: incidence, prevalence, diagnosis, and long-term clinical course of illness. *Scand J Prim Health Care 33* (1):27-32.

6. Nielen MMJ, Boersma-van DAM ME, Schermer TJR (2019) [Incidence and prevelance of health problems in the Dutch general practice in 2017]. Incidentie en prevalentie van gezondheidsproblemen in de Nederlandse huisartsenpraktijk in 2017. From: *NIVEL Zorgregistraties eerste lijn* [internet]. Available from: www.nivel.nl/nl/zorgregistraties-eerste-lijn/incidenties-en-prevalenties. Accessed 2019 Apr 16th. Dutch.

7. Peul WC, van Houwelingen HC, van den Hout WB, Brand R, Eekhof JA, Tans JT, Thomeer RT, Koes BW, Leiden-The Hague

Spine Intervention Prognostic Study G (2007) Surgery versus prolonged conservative treatment for sciatica. *N Engl J Med 356* (22):2245-2256.

8. Konstantinou K, Hider SL, Jordan JL, Lewis M, Dunn KM, Hay EM (2013) The impact of low back-related leg pain on outcomes as compared with low back pain alone: a systematic review of the literature. *Clin J Pain 29* (7):644-654.

9. Kristman VL, Hartvigsen J, Leboeuf-Yde C, Kyvik KO, Cassidy JD (2012) Does radiating spinal pain determine future work disability? A retrospective cohort study of 22,952 Danish twins. *Spine 37* (11):1003-1013.

10. Koes BW, van Tulder MW, Peul WC (2007) Diagnosis and treatment of sciatica. *BMJ 334* (7607):1313-1317. doi:10.1136/bmj.39223.428495.BE

11. Luijsterburg PA, Verhagen AP, Ostelo RW, van Os TA, Peul WC, Koes BW (2007) Effectiveness of conservative treatments for the lumbosacral radicular syndrome: a systematic review. *Eur Spine J 16* (7):881-899.

12. Lewis RA, Williams NH, Sutton AJ, Burton K, Din NU, Matar HE, Hendry M, Phillips CJ, Nafees S, Fitzsimmons D, Rickard I, Wilkinson C (2015) Comparative clinical effectiveness of management strategies for sciatica: systematic review and network meta-analyses. *The spine journal: official journal of the North American Spine Society 15* (6):1461-1477.

13. Fernandez M, Hartvigsen J, Ferreira ML, Refshauge KM, Machado AF, Lemes IR, Maher CG, Ferreira PH (2015) Advice to Stay Active or Structured Exercise in the Management of Sciatica: A Systematic Review and Meta-analysis. *Spine 40* (18):1457-1466.

14. Hayden JA, Cartwright J, van Tulder MW, Malmivaara A (2012) Exercise therapy for chronic low back pain. *Cochrane Database of Systematic Reviews 1* (1):64.

15. Powell KE, Paluch AE, Blair SN (2011) Physical Activity for Health: What Kind? How Much? How Intense? On Top of What? *Annual Review of Public Health 32* (1):349-365.

16. NICE, guideline, [NG59]. (November 2016) Low back pain and sciatica in over 16s: assessment and management. Available from: https://www.nice.org.uk/guidance/NG59/chapter/Recommendations#non-invasive-treatments-for-low-back-pain-and-sciatica. Accessed 2019 March 29th.

17. Fernandez M, Ferreira ML, Refshauge KM, Hartvigsen J, Silva IR, Maher CG, Koes BW, Ferreira PH (2016) Surgery or physical activity in the management of sciatica: a systematic review and meta-analysis. *European Spine Journal. 25* (11):3495-3512.

18. Albert HB, Manniche C (2012) The efficacy of systematic active conservative treatment for patients with severe sciatica: a single-blind, randomized, clinical, controlled trial. *Spine 37* (7):531-542.

19. Luijsterburg PAJ, Verhagen AP, Ostelo R, van den Hoogen H, Peul WC, Avezaat CJJ, Koes BW (2008) Physical therapy plus general practitioners' care versus general practitioners' care alone for sciatica: a randomised clinical trial with a 12-month follow-up. *European Spine Journal 17* (4):509-517.

20. Furlan AD, Pennick V, Bombardier C, van Tulder M (2009) 2009 Updated Method Guidelines for Systematic Reviews in the Cochrane Back Review Group. *Spine 34* (18):1929-1941.

21. Abenhaim L, Rossignol M, Valat JP, Nordin M, Avouac B, Blotman F, Charlot J, Dreiser RL, Legrand E, Rozenberg S, Vautravers P (2000) The role of activity in the therapeutic management of back pain. Report of the International Paris Task Force on Back Pain. *Spine 25* (4 Suppl):1S-33S.

22. Furlan AD, Malmivaara A, Chou R, Maher CG, Deyo RA, Schoene M, Bronfort G, van Tulder MW, Editorial Board of the Cochrane Back NG (2015) 2015 Updated Method Guideline for Systematic Reviews in the Cochrane Back and Neck Group. *Spine*

40 (21):1660-1673.

23. Bakhtiary AH, Safavi-Farokhi Z, Rezasoltani A (2005) Lumbar stabilizing exercises improve activities of daily living in patients with lumbar disc herniation. *Journal of Back and Musculoskeletal Rehabilitation 18* (3-4):55-60.

24. Higgins JPT, Deeks JJ (2011) Chapter 7: Selecting studies and collecting data. In: Higgins JPT, Green S (editors), *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0* (updated March 2011). Available from: www.handbook.cochrane.org. Accessed 2019 September 19th.

25. Higgins JPT, Deeks JJ, Altman DG (2011) Chapter 16: Special topics in statistics. In: Higgins JPT, Green S (editors), *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0* (updated March 2011). Available from: www.handbook.cochrane.org. Accessed 2019 September 19th.

26. Deeks JJ, Higgins JPT, Alstman DG (editors) (2011) Chapter 9: Analysing data and undertaking meta-analyses. In: Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0* (updated March 2011). Available from www.handbook.cochrane.org. Accessed 2019 September 19th.

27. GRADE working group (2004) Grading quality of evidence and strength of recommendations. *BMJ : British Medical Journal 328* (7454):1490-1490.

28. Verwoerd AJH, Luijsterburg PAJ, Koes BW, el Barzouhi A, Verhagen AP (2015) Does kinesiophobia modify the effect of physical therapy on outcome in patients with sciatica in primary care? Subgroup analysis from a randomized controlled trial. *Physical Therapy 2015 Sep;95*(9):1217-1223.

29. Zhang JF, Chen WH (2004) Curative effect of nonoperative therapy for the lumbar disc herniation. *Chinese Journal of Clinical Rehabilitation 8* (11):2006-2007.

30. Zhao Y (2004) Short-term and long-term effect of

movement prescription and recovery education for patients with lumbar intervertebral disc herniation. *Chinese Journal of Clinical Rehabilitation 8* (36):8314-8315.

31. Aras D, Ahmad H (2018) The efficacy of physiotherapy combination technique on pain and functional independence of people with lumbar disc herniation: (physiotherapy combination technique: A conservative treatment for lumbar disc herniation). *Indian Journal of Public Health Research and Development 9* (8):1524-1529.

32. Ding Y, Yang MY (2015) Electroacupuncture assisted by squatting stances for lumbar disc herniation: 128 cases. *World J Acupunct Moxibustion 25* (1):47-50.

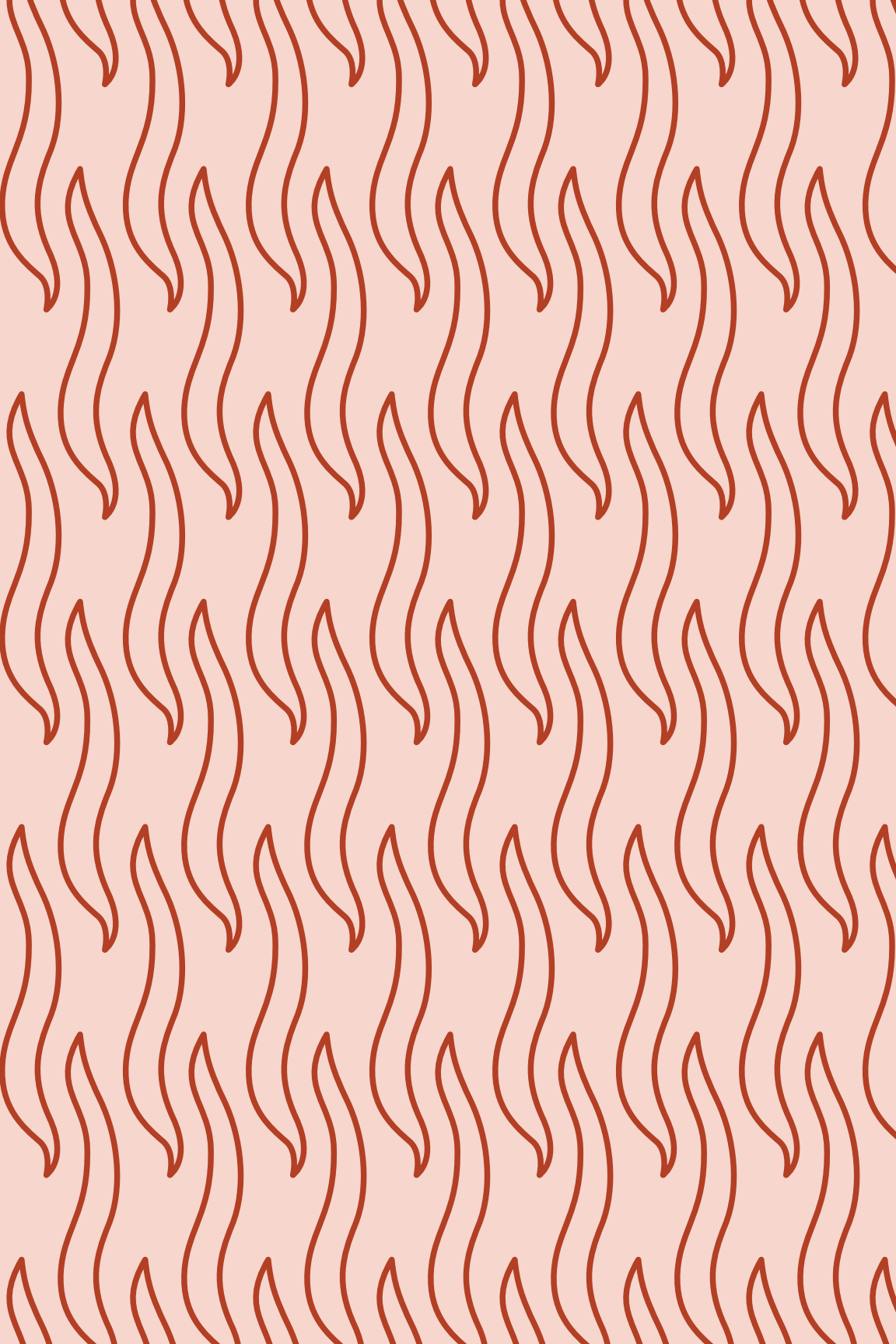
33. Ljunggren AE, Walker L, Weber H, Amundsen T (1992) Manual traction versus isometric exercises in patients with herniated intervertebral lumbar discs. *Physiother Theory Pract 8* (4):207-213.

34. Thackeray A, Fritz JM, Brennan GP, Zaman FM, Willick SE (2010) A pilot study examining the effectiveness of physical therapy as an adjunct to selective nerve root block in the treatment of lumbar radicular pain from disk herniation: a randomized controlled trial. *Physical Therapy 90* (12):1717-1729.

35. Schenk RJ, Jozefczyk C, Kopf A (2003) A randomized trial comparing interventions in patients with lumbar posterior derangement. *J Man Manip Ther 11* (2):95-102.

36. Coxhead CE, Inskip H, Meade TW, North WR, Troup JD (1981) Multicentre trial of physiotherapy in the management of sciatic symptoms. *Lancet 1* (8229):1065-1068.

37. Huber J, Lisiński P, Samborski W, Wytrążek M (2011) The effect of early isometric exercises on clinical and neurophysiological parameters in patients with sciatica: An interventional randomized single-blinded study. *Isokinet Exerc Sci 19* (3):207-214.



C

Methodological studies

7

Current health status strongly influences the Global Perceived Effect scale among sciatica patients who received lumbar disc surgery

Elizabeth N Mutubuki[1], Heleen Kuiper[1,2], Johanna M van Dongen[1], Maurits W van Tulder[1, 3], Steven J Kamper[4], Carmen LA Vleggeert-Lankamp[5], Teddy Oosterhuis[1], Raymond WJG Ostelo[1,6]

[1] Department of Health Sciences, Faculty of Science, Vrije University Amsterdam, Amsterdam Movement Sciences Research Institute, the Netherlands
[2] Centre of Excellence for Rehabilitation Medicine, Brain Centre Rudolf Magnus, University Medical Centre Utrecht, Utrecht University and De Hoogstraat Rehabilitation, the Netherlands
[3] Department of Physiotherapy & Occupational Therapy, Aarhus University Hospital, Aarhus, Denmark
[4] Institute for Musculoskeletal Health, School of Public Health, the University of Sydney, Sydney, Australia
[5] Department of Neurosurgery, Leiden University Medical Centre, Leiden, the Netherlands
[6] Department of Epidemiology and Biostatistics, VU University Medical Center, Amsterdam Movement Sciences, the Netherlands

Under Review

ABSTRACT

Purpose: This study investigated the construct validity of the Global Perceived Effect (GPE) scale for measuring self-perceived recovery in patients with sciatica.
Methods: Data from 169 patients with sciatica were used. Their GPE scores were dichotomized (recovered/non-recovered) and were regressed upon change in leg pain, change in back pain and change in functional status, separately. Then, models were adjusted for current pain or functional status and were re-run for various follow-up durations. Lastly, GPE was regressed upon change in leg pain, change in back pain and change in functional status simultaneously in one model.
Results: The GPE was statistically significantly associated with change in leg pain (OR:1.04;95%CI:1.02-1.05), change in back pain (OR:1.02;95%CI:1.01-1.04), and change in functional status (OR:1.08;95%CI:1.04-1.12). Adding current pain and functional status, respectively, decreased the magnitude of some of the associations and increased all of the models' explained variance. The influence of adjusting for current pain or functional status did not systematically decrease with longer follow-ups durations. Self-perceived recovery had the strongest association with change in functional status (OR:1.25;95%CI:1.10-1.42) when GPE was regressed upon pain and functional status simultaneously.
Conclusion: GPE is not a true transition scale for patients with sciatica, as current health status considerably impacts GPE scores.

Keywords: GPE, current health status, sciatica, change score, postoperative

INTRODUCTION

Sciatica is a common health problem with major social and economic consequences[1,2]. Causes of sciatica include a herniated disc with nerve root compression[3-5]. Sciatica is characterized by radiating pain from the lower back to the lower limbs with neurological deficits. Prevalence estimates of sciatica differ greatly between studies, ranging from 1.6% to 43%[6,7]. In the Netherlands, there were 117,200 new cases of sciatica in 2017[8]. Prognosis of sciatica is favorable for most people with acute sciatica and acute symptoms typically resolve without treatment[3,4,9]. Nonetheless, up to 30% of patients experience pain for longer than one year[10] and about 10% of patients will eventually receive surgery[11]. Along with the human suffering, sciatica is responsible for a large economic burden on society. In the Netherlands, for example, the estimated total annual cost of sciatica is €1.2 billion[12].

Self-perceived recovery is commonly used as an outcome measure to assess the effectiveness of sciatica treatments, such as lumbar disc surgery[13]. Self-perceived recovery is often measured using the Global Perceived Effect (GPE) scale[14-16], which asks patients to indicate how much their condition has improved or deteriorated since a pre-defined time point[17]. The GPE is quick to administer, its results are seemingly simple to interpret, and its test-retest reliability is excellent[17]. The GPE is also used as an external criterion for testing measurement properties of other outcome measures commonly used to assess the effectiveness of sciatica patients, such as pain and disability[14,16,15].

There are some validity concerns about the GPE, with the most important one being whether it is a true “transition scale”. A previous study by Kamper et al. (2010) among patients with musculoskeletal disorders found that GPE scores are greatly influenced by a patient’s current health status and that this effect is more apparent with increasing time periods[17]. This suggests that the GPE does not necessarily measure recovery in terms of a change in health state, but rather a patient’s current health

state, particularly at long-term[17-19] . However, in their study, Kamper et al performed a complete-case analysis, hence their conclusions could have been influenced by selective drop out of patients. Moreover, they only provided crude estimates, meaning that they did not correct for possible confounding factors, such as age and duration of complaints, and included patients who suffered substantial residual complaints six weeks after lumbar disc surgery and not all postoperative sciatica patients[20].

The present study aimed to assess the construct validity of the GPE in postoperative sciatica patients by estimating: 1) the association between the GPE score and change in pain intensity and functional status, 2) the extent to which this association is influenced by a patient's current pain and current functional status, respectively, 3) whether the influence of a patient's current pain and current functional status increased with increasing time periods over which the transitions were measured, and 4) the independent associations of pain and functional status with self-perceived recovery.

METHODS

Study design and population

Data collected during the REALISE (rehabilitation after lumbar disc surgery) study in the Netherlands between 2012-2014 were used for this study[21]. REALISE was a multicenter, randomized controlled trial, which investigated the effectiveness and cost-effectiveness of early rehabilitation compared to no referral for rehabilitation following lumbar disc surgery. Patients who received lumbar disc surgery on a herniated disc confirmed by MRI and signs of nerve root compression corresponding to the level of disc herniation were eligible for the REALISE study. In addition, patients had to be aged between 18-70 years[21]. Exclusion criteria included cauda equine syndrome, neurogenic claudication, previous lumbar disc surgery at the same level on the same side, contraindications for exercise therapy[22]. All patients received postoperative care during their stay in the hospital, regardless of treatment allocation[21]. Intervention

group patients (n=92) received referral for early rehabilitation, control group patients (n=77) received no referral for early rehabilitation. Baseline measurements were done preoperatively, and follow-up measurements at 3, 6, 9, 12 and 26 weeks after surgery[21]. More detailed information about the design and methods of the REALISE study appears elsewhere[22].

Outcome measures

All outcome measures were administered at baseline and after 3, 6, 9, 12, and 26 weeks; only the GPE was not administered at baseline.

Self-perceived recovery

Self-perceived recovery was assessed using the GPE. Patients were asked the following question; "To what extent are your complaints changed when compared with the situation immediately after discharge from the hospital?". Answering options ranged from 1 (completely recovered) to 7 (worst imaginable) and were dichotomized into recovered (including 1; 'completely recovered' and 2; 'much improved') and non-recovered (3; 'slightly improved', 4; 'not changed', 5; 'slightly worsened', 6; 'much worsened', and 7; 'worse than ever')[22]. The GPE demonstrates strong test-retest reliability (ICC was 0.90 to 0.99) for periods up to 24 hours in chronic low back pain patients[17].

Change in leg and low back pain intensity

Pain intensity in leg and back was measured using an 11-point Numeric Rating Scale (NRS), ranging from 0 (no pain) to 10 (worst imaginable pain)[23]. The NRS has good psychometric properties in people with low back pain[24,25] and is one of the core outcome measurement tools for low back pain[26]. The ICC for test-retest reliability was 0.92[27]. Pain scores were transformed to a 0-100 scale to improve comparability with functional status (see below). Change in pain intensity within patients was calculated by subtracting the patients' follow-up NRS scores from their baseline NRS scores. Thus, higher NRS change scores indicate greater improvement.

Change in functional status

Functional status was assessed using the Oswestry Disability Index (ODI)[28,29,21]. The ODI is comprised of 10 questions assessing daily living aspects, such as personal care, lifting, walking, sleeping, sex life, social life, and traveling. ODI scores range from 0 (no difficulty) to 100 (maximal difficulty). The ODI has good psychometric properties in people with low back pain (test-retest ICC was 0.94±0.03 and Cronbach's alpha was 0.88±0.05)[30] and is one of the core outcome measurement tools for low back pain[26]. Change in functional status within patients was calculated by subtracting the patients' follow-up ODI scores from their baseline ODI scores. Thus, higher ODI change scores indicate greater improvement.

Statistical analyses

Missing data were imputed using the Multivariate Imputation by Chained Equations (MICE) procedure[31]. An imputation model was created including all available self-perceived recovery, pain intensity, and functional status values as well as the variables that differed between groups at baseline, were related to the "missingness" of data, and were associated with the outcomes. With this imputation model, 10 complete datasets were created, all of which were analyzed separately. Pooled estimates were calculated according to Rubin's rules[31].

To examine the association between sciatica patients' GPE scores and change in pain intensity and functional status (research question 1), baseline and 26-week follow-up data were used. Three logistic regression analyses were performed, in which GPE (recovered/non-recovered) was the dependent variable and regressed upon; 1) change in leg pain intensity, 2) change in back pain intensity, and 3) change in functional status. Based on work of Waddell et al. (2003), the following variables were identified as potential confounders; pre-operative age (years), sex (male/female), household composition (living with partner/alone/with children< 18 years/with children ≥ 18 years /with parents), level of education (low/middle/high), duration of complaints (weeks), use of pharmaceuticals (yes/no), sick leave (yes/no), type of hernia (sequestered/bulging disc/extraforaminal/other), fear avoidance beliefs (FABQ)[32], pain coping (PCI)[33] and general mental health (SF-12)[34,35]. Confounding was checked by adding the potentially confounding variables to the regression

models; when the coefficient of interest changed by more than 10%, the variable was considered a confounder. To explore whether the association between sciatica patients' GPE scores and change in pain intensity and functional status is influenced by a patient's current pain intensity and current functional status, respectively (research question 2), the patient's 26-week follow-up pain intensity score or functional status score was added to the aforementioned regression models as a current score. Odds Ratios (ORs) and R^2 values from adjusted and unadjusted regression models were compared. To explore whether the influence of a patient's current pain intensity and current functional status increased with increasing follow-up periods (research question 3), all of the aforementioned analyses were repeated with 3, 6, 9, and 12 weeks follow-up pain intensity and functional status scores and we compared whether the difference between the ORs and R^2 values from adjusted and unadjusted analyses systematically in- or decreased with increasing follow-up durations. To explore the independent associations of pain and functional status with GPE (research question 4), a multivariable logistic regression model was constructed, in which self-perceived recovery (recovered/non-recovered) was the dependent variable and regressed upon all 26-week variables simultaneously i.e. change in leg pain, change in back pain and change in function. Analyses were conducted in IBM SPSS version 25 and statistical significance was set at $p \leq 0.05$.

RESULTS

Study population

A total of 169 sciatica patients participated in the REALISE study, 58% were female (n=98). Of the patients, 74% (n=125) lived with a partner and 23% (n=38) were unemployed. A middle education level was most common (48%; n=82), followed by a high level (30%; n=50) and a low level (22%; n=37). The most common type of disc herniation was bulging disc (51%; n=86), followed by sequestered disc herniation (39%; n=66), then other types of disc herniation (9%; n=16) and lastly extraforaminal disc

herniation (1%; n=1). The most common levels of disc herniation were L5-S1 (45%; n=77) and L4-L5 (43%; n=73) (Table 1). During the course of the study, 19 patients were lost to follow-up. Complete data on all follow-up measurement points were available from 88% of patients (n=148).

Table 1. Patient characteristics

Characteristic	**Sample distribution % (n)**
Sex	
Male	42 (71)
Female	58 (98)
Household composition	
Living alone: Yes	13,6 (23)
No	86,4 (146)
Living with a partner	*74 (125)*
Living without a partner	*26 (44)*
With children < 18 years: Yes	*39,6 (67)*
With children < 18 years: No	*60,4 (102)*
With children ≥ 18 years: Yes	*18,9 (32)*
With children ≥ 18 years: No	*81,1 (137)*
Living with parents: Yes	*4,7 (8)*
No	*95,3 (161)*
Education	
Low	22 (37)
Middle	48 (82)
High	30 (50)
Employment	
Unemployed	22,5 (38)
Employed	77,5 (131)
Level of herniation	
L2-3	2 (3)
L3-4	8 (14)
L4-5	43 (73)
L5-S1	45 (77)
L5-6	2 (3)
Type of herniation	
Sequester	39,1(66)
Bulging disc	50,9 (86)
Extraforaminal	0,6 (1)
Other	9,4 (16)

(Oosterhuis et al., 2017) N=169.

Association between GPE and change in pain and functional status

At 26 weeks, the GPE was statistically significantly associated with change in leg pain (OR: 1.04; 95%CI:1.02 to 1.05), change in back pain (OR: 1.02; 95%CI: 1.01 to 1.04), and change in functional status (OR: 1.08; 95%CI: 1.04 to 1.12). Thus, for every point (100-point scale) improvement in leg pain, back pain, and functional status from baseline to 26-week follow-up, the odds of rating oneself as being successfully recovered was 1.04, 1.02, and 1.08, respectively. The R^2 values for the association between GPE and change in leg pain, change in back pain, and change in functional status were 31%, 13%, and 42%, respectively (Table 2).

Table 2. Association between self-perceived recovery (GPE), and change in pain intensity (leg and back) and functional status at 26 weeks (with and without accounting for current score)

		Beta	SE	R^2	P-value	Odds ratio	Lower bound	Upper bound
	Unadjusted for Current Score							
GPE 26 weeks	*Change in pain intensity in leg (Range: 0-100)	**0.04**	**0.01**	**0.31**	**≤0.001**	**1.04**	**1.02**	**1.05**
	**Change in pain intensity in back (Range: 0-100)	**0.02**	**0.01**	**0.13**	**0.003**	**1.02**	**1.01**	**1.04**
	***Change in functional status (Range: 0-100)	**0.07**	**0.02**	**0.42**	**≤0.001**	**1.08**	**1.04**	**1.12**
	Adjusted for Current Score							
	*Change in pain intensity in leg (Range: 0-100)	**0.01**	**0.01**	**0.63**	**0.579**	**1.01**	**0.98**	**1.04**
	**Change in pain intensity in back (Range: 0-100)	**0.01**	**0.01**	**0.52**	**0.430**	**1.01**	**0.99**	**1.03**
	***Change in functional status (Range: 0-100)	**0.14**	**0.05**	**0.61**	**0.007**	**1.15**	**1.05**	**1.27**

**No confounding factors; none of the confounders changed the regression co-efficient by more than 10%*

***No confounding factors; none of the confounders changed the regression co-efficient by more than 10%*

****Adjusted for the following confounders pain coping (PCI) and amount of pills taken again the pain*

With exception to the p-values, all other numbers have been rounded to two decimal places

Table 3. Association between self-perceived recovery and change in pain intensity and functional status at 3, 6, 9, 12, 26 weeks

		Beta	SE	R2	P-value	Odds ratio	Lower bound	Upper bound
GPE 3 weeks	**Unadjusted for Current Score**							
	*Change in pain intensity in leg (Range: 0-100)	0.05	0.01	0.42	≤ 0.001	1.05	1.03	1.06
	**Change in pain intensity in back (Range: 0-100)	0.02	0.05	0.09	0.002	1.02	1.01	1.03
	***Change in functional status scores (Range: 0-100)	0.08	0.01	0.41	≤ 0.001	1.08	1.05	1.11
	Adjusted for Current Score							
	*Change in pain intensity in leg (Range: 0-100)	0.02	0.01	0.51	0.092	1.02	1.00	1.04
	**Change in pain intensity in back (Range: 0-100)	0.02	0.06	0.34	0.006	1.02	1.01	1.03
	***Change in functional status scores (Range: 0-100)	0.04	0.02	0.49	0.013	1.04	1.01	1.08
GPE 6 weeks	**Unadjusted for Current Score**							
	*Change in pain intensity in leg (Range: 0-100)	0.04	0.01	0.32	≤ 0.001	1.04	1.03	1.06
	**Change in pain intensity in back (Range: 0-100)	0.03	0.01	0.17	≤ 0.001	1.03	1.01	1.04
	***Change in functional status scores (Range: 0-100)	0.06	0.01	0.28	≤ 0.001	1.06	1.03	1.09
	Adjusted for Current Score							
	*Change in pain intensity in leg (Range: 0-100)	0.0	0.01	0.43	0.119	1.02	1.00	1.04
	**Change in pain intensity in back (Range: 0-100)	0.01	0.01	0.31	0.219	1.01	0.99	1.03
	***Change in functional status scores (Range: 0-100)	0.02	0.02	0.54	0.405	1.02	0.98	1.05
GPE 9 weeks	**Unadjusted for Current Score**							
	*Change in pain intensity in leg (Range: 0-100)	0.05	0.01	0.40	≤ 0.001	1.05	1.03	1.06
	**Change in pain intensity in back (Range: 0-100)	0.04	0.01	0.28	≤ 0.001	1.04	1.02	1.05
	***Change in functional status scores (Range: 0-100)	0.07	0.02	0.39	≤ 0.001	1.07	1.04	1.10
	Adjusted for Current Score							
	*Change in pain intensity in leg (Range: 0-100)	0.02	0.01	0.65	0.264	1.02	0.99	1.04
	**Change in pain intensity in back (Range: 0-100)	0.03	0.01	0.52	0.009	1.03	1.01	1.05
	***Change in functional status scores (Range: 0-100)	0.02	0.22	0.65	0.483	1.02	9.73	1.06

Continued - Table 3. Association between self-perceived recovery and change in pain intensity and functional status at 3, 6, 9, 12, 26 weeks

	Unadjusted for Current Score							
GPE 12 weeks	*Change in pain intensity in leg (Range: 0-100)	**0.05**	**0.01**	**0.39**	**≤ 0.001**	**1.05**	**1.03**	**1.07**
	**Change in pain intensity in back (Range: 0-100)	**0.03**	**0.01**	**0.18**	**≤ 0.001**	**1.03**	**1.02**	**1.04**
	***Change in functional status scores (Range: 0-100)	**0.10**	**0.02**	**0.49**	**≤ 0.001**	**1.10**	**1.06**	**1.15**
	Adjusted for Current Score							
	*Change in pain intensity in leg (Range: 0-100)	**0.01**	**0.02**	**0.65**	**0.379**	**1.01**	**0.98**	**1.04**
	**Change in pain intensity in back (Range: 0-100)	**0.02**	**0.01**	**0.37**	**0.112**	**1.02**	**1.00**	**1.03**
	***Change in functional status scores (Range: 0-100)	**0.04**	**0.03**	**0.69**	**0.089**	**1.05**	**0.99**	**1.09**
	Unadjusted for Current Score							
GPE 26 weeks	*Change in pain intensity in leg (Range: 0-100)	**0.04**	**0.01**	**0.31**	**≤0.001**	**1.04**	**1.02**	**1.05**
	**Change in pain intensity in back (Range: 0-100)	**0.02**	**0.01**	**0.13**	**0.003**	**1.02**	**1.01**	**1.04**
	***Change in functional status (Range: 0-100)	**0.07**	**0.02**	**0.42**	**≤ 0.001**	**1.08**	**1.04**	**1.12**
	Adjusted for Current Score							
	*Change in pain intensity in leg (Range: 0-100)	**0.01**	**0.01**	**0.63**	**0.579**	**1.01**	**0.98**	**1.04**
	**Change in pain intensity in back (Range: 0-100)	**0.01**	**0.01**	**0.52**	**0.430**	**1.01**	**0.99**	**1.03**
	***Change in functional status (Range: 0-100)	**0.14**	**0.05**	**0.61**	**0.007**	**1.15**	**1.05**	**1.27**

** No confounding factors; none of the confounders changed the regression co-efficient by more than 10%*
*** No confounding factors; none of the confounders changed the regression co-efficient by more than 10%*
**** Adjusted for the following confounders pain coping (PCI) and amount of pills taken again the pain*
With exception to the p-values, all other numbers have been rounded to two decimal places

Influence of current health status

After accounting for the patients' current pain or functional status score, ORs became smaller for the associations between GPE and change in back and leg pain and both associations were no longer statistically significant (Table 2). For functional status, the association increased and remained statistically significant (OR: 1.15; 95%CI: 1.05 to 1.27). Moreover, the R^2 values increased for all associations, meaning that the models' explained variance increased after accounting for the patients' current pain or functional status score (Table 2).

Influence of a patient's current health status with increasing time periods

The difference between the ORs and R^2 values from adjusted and unadjusted analyses did not systematically in- or decrease with increasing follow-up periods (Table 3). This indicates that the impact of current pain and current functional status on GPE did not become systematically more or less pronounced the longer or shorter the recall period.

The independent associations of pain and functional status with GPE

When regressing GPE upon change in back pain, change in leg pain, and change in functional status simultaneously, change in functional status was found to have the stronger association with GPE, followed by change in leg pain and then change in back pain. After accounting for the patients' current pain or functional status score, change in functional status was still found to have the stronger association with GPE (OR: 1.25; 95%CI 1.10 to 1.42) (Table 4).

Table 4.a. Contribution of pain and functionality on GPE, unadjusted for current score (pain, function)

	B (SE)	Wald	P-value	95%CI		
				Odds Ratio	Lower bound	Upper bound
Change in pain intensity in leg (Range: 0-100)	0.03 (0.01)	8.63	0.060	1.03	1.01	1.06
*Change in pain intensity in back (Range scores 0-100	0.01 (0.01)	0.35	0.557	1.01	0.99	1.03
Change in functional status scores (Range: 0-100)	0.07 (0.03)	8.40	0.009	1.07	1.02	1.13
Pain coping (PCI)	-0.19 (0.23)	0.67	0.416	0.83	0.53	1.31
Average amount of tablets taken per day against the pain	-0.21 (0.07)	8.33	0.040	0.81	0.70	0.93
Constant (i.e. intercept)	-1.04 (1.47)	0.51	0.479			

R2 is 0.605
With exception to the p-values, all other numbers have been rounded to two decimal places

Table 4.b. Contribution of pain and functionality on GPE, adjusted for current score (pain, function)

	B (SE)	Wald	P-value	95%CI		
				Odds Ratio	Lower bound	Upper bound
Change in pain intensity in leg (Range: 0-100)	-0.04 (0.03)	0.02	0.880	1.00	0.95	1.05
Pain intensity in leg (Range: 0-100) (current score)	-0.04 (0.03)	2.11	0.146	0.96	0.90	1.02
*Change in pain intensity in back (Range scores 0-100)	0.03 (0.02)	1.95	0.163	1.03	0.99	1.07
Pain intensity in back (Range: 0-100) (current score)	-0.01 (0.02)	0.14	0.708	0.99	0.95	1.04
Change in functional status scores (Range: 0-100)	0.22 (0.06)	12.24	≤ 0.001	1.25	1.10	1.42
Functional status (Range: 0-100) (current score)	-0.43 (0.13)	10.55	0.001	0.65	0.50	0.84
Pain coping (PCI)	0.67 (0.37)	3.26	0.071	1.95	0.94	4.04
Average amount of tablets taken per day against the pain	-0.18 (0.11)	2.66	0.103	0.84	0.68	1.04
Constant (i.e. intercept)	1.77 (2.29)	0.60	0.440			

R2 is 0.843
With exception to the p-values, all other numbers have been rounded to two decimal places

DISCUSSION

The present study assessed the construct validity of the GPE among sciatica patients who received lumbar disc surgery. The GPE was found to be statistically significantly associated with change in leg pain, change in back pain, and change in functional status. To illustrate, the odds of rating oneself as being successfully recovered with a 1 point improvement on a 0-100 scale in leg pain, back pain and functional status was found to be 1.04, 1.02, and 1.08, respectively. Interpreting these ORs in terms of clinical relevance (defined as a 10-point increase on the ODI [range: 0-100] and a 25-point increase on the NRS [range: 0-100][25]), the odds of rating oneself as being successfully recovered with a clinically meaningful change was 2.16 for functional status and 2.67 and 1.64 for leg and back pain, respectively. However, when the patients' current pain or functional status scores were added to the regression equations, all pain associations decreased in size, while all of the models' explained variance improved. This shows that a patient's current health status influences whether they consider themselves recovered or not on the GPE scale. Further, the influence of adding the patients' current pain or functional status scores to the regression equations did not systematically increase with increasing follow-up durations. Thus, a patient's current health status has a considerable impact on their GPE score regardless of follow-up duration, which means the GPE scale is not a true "transition scale", even at short follow-up durations. Finally, change in functional status had the strongest association with GPE. This means that postoperative sciatica patients were more likely to view themselves as recovered when they experienced improved function, as opposed to improved back pain or leg pain.

Comparison with literature

The current findings indicate that patients take into account their current health status when scoring the GPE. This is in line with the findings of Schmitt and Fabio (2005) who assessed the validity of retrospective measures of change in patients with upper extremity musculoskeletal problems, and concluded they were not accurate measures of change over time[36]. Kamper

et al (2010) also found that GPE scores were influenced by a patient's current status and came to the same conclusion[17]. Both studies observed the similar results, in different patient populations. That is, Schmitt and Fabio (2005) included patients with upper extremity musculoskeletal problems, whereas Kamper et al (2010) only included sciatica patients who suffered substantial residual complaints six weeks after lumbar disc surgery. We included postoperative sciatica patients in general in the present study.

Both Schmitt and Fabio (2005) and Kamper et al (2010) concluded that GPE scales performed adequately over short transition periods and were less accurate with longer recall periods. Our findings do not support either of these propositions, and indicate that GPE scales do not perform adequately as a transition scale, regardless of transition period. Further research into an alternative is warranted, as the GPE does not seem to be a true transition scale.

The present study also found that change in functional status had the strongest association with GPE, suggesting that patients base their self-perceived recovery score more on their change in functional status than on pain intensity. This is in line with a recent finding from our research group, which showed that disability had stronger longitudinal associations with health-related quality of life and costs than pain[37].These finding suggest that it is not a patient's level of pain intensity, but the way pain influences their daily activities that most influences a person's self-perceived health status.

Strengths and Limitations

Strengths of the present study include that it is one of the first studies to assess the construct validity of the GPE amongst postoperative sciatica patients. A second strength of this study is that we controlled for confounders, which may have influenced the results and conclusions of previous studies. The use of multiple imputation to deal with missing data, thereby avoiding a complete-case analysis, increased power and reduced the risk of bias due to selective drop-out of patients[38], is another strength. A potential limitation of this study is generalizability of these results to other populations beyond postoperative sciatica patients. On the other hand, it adds to the existing evidence

regarding the construct validity of the GPE as we specifically focused on another patient group in the musculoskeletal domain. The study population consists of RCT participants, which may be slightly different to sciatica patients in general. However, as no differences in clinical outcomes were found between interventions groups at all measurement points[21] and similar results were found in other patient populations, we do not expect this to have biased our results substantially.

Implications for clinical practice and research

The GPE is highly influenced by a patient's current health status and is therefore not a "true transition scale" when assessing recovery in postoperative sciatica patients. Clinicians interested in centering treatment decisions on change status should be aware that the GPE is not a true transition scale. Clinicians could ensure that they note the baseline scores (i.e. pain, function) and follow-up scores of their patients and then calculate the corresponding change scores themselves. Even though the GPE is a simple and easy to use tool in clinical trials, researchers should bear in mind that it does not seem to be a gold standard for measuring transitions in health among patient populations.

Conclusions

GPE is highly influenced by a patient's current health status, even at short-term follow-up, and can therefore not be considered a true transition scale. The current study findings also suggest that GPE is influenced more by functional status than by pain. Further research into an alternative scale for the GPE is warranted.

DISCLOSURES

Funding The REALISE study was funded by the Netherlands Organization for Health Research and Development (REALISE-171102010).

Ethics approval Ethical approval for the REALISE study was obtained from the Medical Ethics Review Board of the VU University Medical Centre in Amsterdam (registration number NL35897.029.11). Local research governance was obtained from all participating hospitals and primary care facilities.

Code availability Codes are available upon reasonable request.

REFERENCES

1. Vlaeyen JWS MC, Wiech K, van Zundert J, Beraldo Meloto C, Diatchenko L, Michele C, Battié MC, Goossens M, Koes B, Linton SJ (2018) Low back pain. Nature Reviews Disease Primers 4 (52).

2. Maslak JP, Jenkins TJ, Weiner JA, Kannan AS, McCarthy MH, Hsu WK, Patel AA (2019) Burden of Sciatica on US Medicare Recipients. J Am Acad Orthop Surg. doi:10.5435/JAAOS-D-19-00174

3. Koes BW, van Tulder MW, Peul WC (2007) Diagnosis and treatment of sciatica. BMJ 334 (7607):1313-1317. doi:10.1136/bmj.39223.428495.BE

4. Ropper AH, Zafonte RD (2015) Sciatica. N Engl J Med 372 (13):1240-1248. doi:10.1056/NEJMra1410151

5. Valat JP, Genevay S, Marty M, Rozenberg S, Koes B (2010) Sciatica. Best Pract Res Clin Rheumatol 24 (2):241-252. doi:10.1016/j.berh.2009.11.005

6. Videman T, Nurminen T, Kuorinka I, Vanharanta H, Troup JD (1984) Low-back pain in nurses and some loading factors of work. Spine 9 (4):400-404.

7. Deyo RA, Tsui-Wu YJ (1987) Descriptive epidemiology of low-back pain and its related medical care in the United States. Spine (Phila Pa 1976) 12 (3):264-268. doi:10.1097/00007632-198704000-00013

8. Nivel Zorgregistraties eerste lijn (2018) Nivel Zorgregistraties.

9. Vroomen PC, de Krom MC, Wilmink JT, Kester AD, Knottnerus JA Lack of effectiveness of bed rest for sciatica. N Engl J Med, 340 (1999):418-423.

10. Vroomen PC, de Krom MC, Slofstra PD, Knottnerus JA (2000) Conservative treatment of sciatica: a systematic review. J Spinal

Disord 13 (6):463-469. doi:10.1097/00002517-200012000-00001

11. van Beek ELK, van Schooten G, Vlieger EJ (2010) Reduceren van praktijkvariatie: budgettaire effecten van scherpere indicatiestelling. Reduceren van praktijkvariatie: budgettaire effecten van scherpere indicatiestelling. Breukelen: Plexus.

12. Health council of the Netherlands Management of Lumbosacral Radicular Syndrome (sciatica). Health council of the Netherlands, The Hague.

13. Oosterhuis T, Costa LO, Maher CG, de Vet HC, van Tulder MW, Ostelo RW (2014) Rehabilitation after lumbar disc surgery. Cochrane Database Syst Rev (3):CD003007. doi:10.1002/14651858.CD003007.pub3

14. Pengel LH, Refshauge KM, Maher CG (2004) Responsiveness of pain, disability, and physical impairment outcomes in patients with low back pain. Spine (Phila Pa 1976) 29 (8):879-883. doi:10.1097/00007632-200404150-00011

15. van der Roer N, Ostelo RW, Bekkering GE, van Tulder MW, de Vet HC (2006) Minimal clinically important change for pain intensity, functional status, and general health status in patients with nonspecific low back pain. Spine (Phila Pa 1976) 31 (5):578-582. doi:10.1097/01.brs.0000201293.57439.47

16. Stewart M, Maher CG, Refshauge KM, Bogduk N, Nicholas M (2007) Responsiveness of pain and disability measures for chronic whiplash. Spine (Phila Pa 1976) 32 (5):580-585. doi:10.1097/01.brs.0000256380.71056.6d

17. Kamper SJ, Ostelo RW, Knol DL, Maher CG, de Vet HC, Hancock MJ (2010) Global Perceived Effect scales provided reliable assessments of health transition in people with musculoskeletal disorders, but ratings are strongly influenced by current status. J Clin Epidemiol 63 (7):760-766 e761. doi:10.1016/j.jclinepi.2009.09.009

18. Norman GR, Stratford P, Regehr G (1997) Methodological problems in the retrospective computation of responsiveness to change: the lesson of Cronbach. J Clin Epidemiol 50 (8):869-879. doi:10.1016/s0895-4356(97)00097-8

19. Guyatt GH, Norman GR, Juniper EF, Griffith LE (2002) A critical look at transition ratings. Journal of clinical epidemiology 55 (9):900-908.

20. Ostelo RW, de Vet HC, Berfelo MW, Kerckhoffs MR, Vlaeyen JW, Wolters PM, van den Brandt PA (2003) Effectiveness of behavioral graded activity after first-time lumbar disc surgery: short term results of a randomized controlled trial. Eur Spine J 12 (6):637-644. doi:10.1007/s00586-003-0560-9

21. Oosterhuis T, Ostelo RW, van Dongen JM, Peul WC, de Boer MR, Bosmans JE, Vleggeert-Lankamp CL, Arts MP, van Tulder MW (2017) Early rehabilitation after lumbar disc surgery is not effective or cost-effective compared to no referral: a randomised trial and economic evaluation. Journal of Physiotherapy

22. Oosterhuis T, van Tulder M, Peul W, Bosmans J, Vleggeert-Lankamp C, Smakman L, Arts M, Ostelo R (2013) Effectiveness and cost-effectiveness of rehabilitation after lumbar disc surgery (REALISE): design of a randomised controlled trial. BMC Musculoskelet Disord 14 (1):124. doi:10.1186/1471-2474-14-124

23. Downie WW, Leatham PA, Rhind VM, Wright V, Branco JA, Anderson JA (1978) Studies with pain rating scales. Ann Rheum Dis 37 (4):378-381. doi:10.1136/ard.37.4.378

24. Childs JD, Piva SR, Fritz JM (2005) Responsiveness of the numeric pain rating scale in patients with low back pain. Spine (Phila Pa 1976) 30 (11):1331-1334. doi:10.1097/01.brs.0000164099.92112.29

25. Ostelo RW, de Vet HC (2005) Clinically important outcomes in low back pain. Best Pract Res Clin Rheumatol 19 (4):593-607. doi:10.1016/j.berh.2005.03.003

26. Chiarotto A, Boers M, Deyo RA, Buchbinder R, Corbin TP, Costa LOP, Foster NE, Grotle M, Koes BW, Kovacs FM, Lin CC, Maher CG, Pearson AM, Peul WC, Schoene ML, Turk DC, van Tulder MW, Terwee CB, Ostelo RW (2018) Core outcome measurement instruments for clinical trials in nonspecific low back pain. Pain 159 (3):481-495. doi:10.1097/j.pain.0000000000001117

27. Chiarotto A, Maxwell LJ, Ostelo RW, Boers M, Tugwell P, Terwee CB (2019) Measurement Properties of Visual Analogue Scale, Numeric Rating Scale, and Pain Severity Subscale of the Brief Pain Inventory in Patients With Low Back Pain: A Systematic Review. J Pain 20 (3):245-263. doi:10.1016/j.jpain.2018.07.009

28. Fairbank JC, Pynsent PB (2000) The Oswestry Disability Index. Spine (Phila Pa 1976) 25 (22):2940-2952; discussion 2952. doi:10.1097/00007632-200011150-00017

29. Little DG, MacDonald D (1994) The use of the percentage change in Oswestry Disability Index score as an outcome measure in lumbar spinal surgery. Spine 19 (19):2139-2142.

30. Sheahan PJ, Nelson-Wong EJ, Fischer SL (2015) A review of culturally adapted versions of the Oswestry Disability Index: the adaptation process, construct validity, test-retest reliability and internal consistency. Disabil Rehabil 37 (25):2367-2374. doi:10.3109/09638288.2015.1019647

31. White IR, Royston P, Wood AM (2011) Multiple imputation using chained equations: issues and guidance for practice. Statistics in medicine 30 (4):377-399.

32. Waddell G, Newton M, Henderson I, Somerville D, Main CJ (1993) A Fear-Avoidance Beliefs Questionnaire (FABQ) and the role of fear-avoidance beliefs in chronic low back pain and disability. Pain 52:157-168.

33. Kraaimaat FW, Evers AW (2003) Pain-coping strategies in chronic pain patients: psychometric characteristics of the pain-coping inventory (PCI). Int J Behav Med 10:343-363.

34. Waddell G, Burton A, Main CJ (2003) Screening to identify people at risk of long-term incapacity for work. London: RSM.

35. Xuemei L, George ML, Kakouras I, Edwards CL, Pietrobon R, Richardson W, Hey L (2003) Reliability, Validity, and Responsiveness of the Short Form 12-Item Survey (SF-12) in Patients With Back Pain. Spine 28 (15):1739-1745.

36. Schmitt J, Di Fabio RP (2005) The validity of prospective and retrospective global change criterion measures. Arch Phys Med Rehabil 86 (12):2270-2276. doi:10.1016/j.apmr.2005.07.290

37. Mutubuki EN, Beljon Y, Maas ET, Huygen FJPM, Ostelo RWJG, van Tulder MW, van Dongen JM (2019) The longitudinal relationships between pain severity and disability versus health-related quality of life and costs among chronic low back pain patients. Quality of Life Research:1-13.

38. Burton A BL, Bryan S (2007) Cost-effectiveness in clinical trials: using multiple imputation to deal with incomplete cost data. Clinical Trials 4 (2):154-161

8

The statistical approach in trial-based economic evaluations matters: get your statistics together!

***Elizabeth N. Mutubuki,[1] *Mohamed El Alili,[2] Judith E. Bosmans,[2] Teddy Oosterhuis,[2] Frank Snoek,[3] Raymond W. J. G. Ostelo,[1,4] Maurits W. van Tulder,[2,5] Johanna M. van Dongen,[1,2]**
***Contributed equally**

[1] Department of Health Sciences, Faculty of Science, VU Amsterdam, Amsterdam Movement Sciences Research Institute, the Netherlands
[2] Department of Health Sciences, Faculty of Science, VU Amsterdam, Amsterdam Public Health Research Institute, the Netherlands
[3] Department of Medical Psychology, Amsterdam UMC, VU Amsterdam, Amsterdam, the Netherlands
[4] Department of Epidemiology and Biostatistics, Amsterdam UMC, Location VUmc, Amsterdam Movement Sciences Research Institute, the Netherlands
[5] Department of Physiotherapy & Occupational Therapy, Aarhus University Hospital, Aarhus, Denmark

Under Review

ABSTRACT

Background: Baseline imbalances, skewed costs, the correlation between costs and effects, and missing data are statistical challenges that are often not adequately accounted for in the analysis of cost-effectiveness data. This study aims to evaluate whether or not accounting for these statistical challenges in trial-based economic evaluations has an impact on the results.
Methods: Data from two trial-based economic evaluations, the REALISE and HypoAware studies, were used. In total, 14 full cost-effectiveness analyses were performed per study, in which all of the statistical challenges in trial-based economic evaluations were taken into account step-by-step. Statistical approaches were compared in terms of the resulting cost and effect differences, ICERs, and probabilities of cost-effectiveness.
Results: In the REALISE study and HypoAware study, the ICER ranged from 636,744€/QALY and 90,989€/QALY when ignoring all statistical challenges to -7,502€/QALY and 46,592€/QALY when accounting for all statistical challenges, respectively. The probabilities of the intervention being cost-effective at 0€/QALY gained were 0.67 and 0.59 when ignoring all statistical challenges, and 0.54 and 0.27 when all of the statistical challenges were taken into account for the REALISE study and HypoAware study, respectively.
Conclusion: The current findings indicate that not taking into account baseline imbalances, skewed costs, correlated costs and effects, and missing data in trial-based economic evaluations may notably impact results. Therefore, when conducting trial-based economic evaluations, it is important to first check the data to identify statistical challenges that need to be adequately accounted for in the analysis.

Key words: cost-benefit analysis, methodology, clinical trial, multiple imputation, baseline imbalances, skewed costs

INTRODUCTION

Economic evaluations aim to inform resource allocation decisions in healthcare by evaluating whether the additional health benefits of an intervention justify its additional costs[1]. In many countries, economic evaluations are increasingly being accepted as a formal decision criterion for the reimbursement of pharmaceuticals and other healthcare technologies[2, 3]. Consequently, there is growing interest in economic evaluations of healthcare interventions.

Although great improvements in the conduct and reporting of economic evaluations along clinical trials have been made in previous years[4], literature shows that the quality of the applied statistical methods is typically far from optimal[4-6]. Often, baseline imbalances, the skewed nature of cost data, and the correlation between costs and effects are not adequately accounted for[7, 8]. Additionally, missing data are frequently handled using "naïve" imputation methods, such as mean imputation and last observation carried forward[7-11]. Failure to appropriately account for these statistical issues when analyzing trial-based economic evaluations is of great concern, because use of inadequate statistical methods may lead to biased results, and consequently invalid decisions resulting in a potential waste of scarce resources [12].

In recent years, various studies have been performed on how to deal with baseline imbalances, skewed costs, correlated costs and effects, and missing data in trial-based economic evaluations[8-10, 13-18]. A more detailed description of these statistical issues and how to deal with them can be found in Box 1[19]. Although previous studies have investigated how to deal with these statistical challenges in trial-based economic evaluations separately, the impact of simultaneously accounting for these statistical challenges on the results of trial-based economic evaluations has not yet been explored. The current study aims to address this gap in knowledge, by analyzing data from two previous trial-based economic evaluations, the REALISE and HypoAware studies, whilst step-by-step taking

into account the aforementioned statistical challenges in the analysis of the data.

Box 1: Statistical challenges in trial-based economic evaluations

1) Baseline imbalances

It is commonly assumed that the random allocation of participants in trial-based economic evaluations ensures that observed and non-observed characteristics are well-balanced across study groups. Nevertheless, some between-group differences in baseline values and/or important prognostic factors regularly occur[20]. Failure to account for such baseline imbalances will likely lead to biased results[20, 21]. Various methods have been suggested to account for baseline imbalances in trial-based economic evaluations, including mean difference adjustment, regression-based adjustment, and matching methods[14, 16, 20-23].

2) Skewed costs

Costs are generally right-skewed as there are relatively few participants with (very) high costs and it is impossible to incur negative costs. Consequently, the assumption of normality of standard parametric tests, such as t-tests and linear regression analyses, is violated[4, 12]. Although the normality assumption is violated by the skewed nature of costs, if the sample size is large enough the central limit theorem ensures that sample means will be normally distributed and standard parametric statistical methods may be used[24]. Log-transformations and standard non-parametric tests (e.g., Mann-Whitney U) are unsuitable for trial-based economic evaluations, since both methods fail to provide an estimate of the mean difference in costs, whereas this is required by decision-makers to allow for estimations of the total budget impact of a new intervention[1, 8, 25, 26]. Suggested methods that are suitable to account for the skewness of costs, while simultaneously comparing mean costs are non-parametric bootstrapping and generalized linear models assuming distributions that fit the data best (e.g. Gamma, Log-Normal and Inverse-Gaussian)[1, 4, 8, 12, 15, 25-29].

3) Correlated costs and effects

Costs and effects are typically correlated and hence their correlation should be accounted for[30-32]. Proposed methods include non-parametric bootstrapping, which can account for the correlation between costs and effects by resampling them in pairs[10], and seemingly unrelated regression (SUR), in which two separate regression models are specified simultaneously; i.e. one for costs and one for effects. In SUR, the correlation between costs and effects is accounted for through correlated error terms[33, 34].

4) Missing data

In clinical trials, missing data are common[1, 25, 26, 35]. This is of great concern for trial-based economic evaluations, because total costs are calculated as the sum of several cost components measured at different time points. If one resource use item or one time point is missing, total costs will also be missing[1, 25, 26]. According to Rubin[35], missing data can be classified in three mechanisms. First, if missing values are not dependent on any observed or unobserved variable, data are said to be missing completely at random (MCAR). Second, when missing values are related to one or more observed variables, but not the missing value itself, data are said to be missing at random (MAR). Third, when missing data depends on the missing values itself, data are said to be missing not at random (MNAR). In case of MAR and MNAR, bias may be introduced when analyses are restricted to complete cases or when naïve imputation methods, such as mean imputation or last observation carried forward, are used[1, 25, 26, 36]. Only, when missing data can validly assumed to be MCAR or the proportion of missing data is low (i.e. <5%), naïve imputation methods may be used[37, 38]. For all other situations, "naïve" imputation methods are discouraged, because - amongst others - they do not account for the uncertainty related to imputing missing values[9, 11, 39-41]. More advanced methods to account for missing data assuming MAR in trial-based economic evaluations include multiple imputation and statistical models with maximum likelihood estimation[9, 11, 39, 40].

METHODS

Data

To evaluate the impact of whether or not accounting for baseline imbalances, skewed costs, correlated costs and effects, and missing data in trial-based economic evaluations, empirical data from two previously published trial-based economic evaluations were used, the REALISE and HypoAware study.

REALISE study

In the Rehabilitation After Lumbar disc Surgery (REALISE) study, early rehabilitation after lumbar disc surgery was compared to no referral after lumbar disc surgery among 169 participants (intervention group: n = 92; control group: n = 77). Resource use was measured from a societal perspective at 6, 12 and 26 weeks follow-up using cost questionnaires[42]. Resource use was valued using Dutch standard costs[43]. Utility values were based on the EuroQol (EQ-5D-3L), which was administered at baseline and 3, 6, 9, 12 and 26 weeks follow-up[42]. Utility values were estimated using the Dutch tariff for the EQ-5D-3L[44]. Quality-adjusted life years (QALYs) were calculated using linear interpolation between measurement points.

HypoAware study

In the HypoAware study, the HypoAware intervention (a blended, group and online psycho-educational intervention based on the evidence-based Blood Glucose Awareness Training) was compared to usual care among 137 participants (intervention group: n = 71; control group: n = 66)[45]. Resource use was measured from a societal perspective at 2, 4, and 6 months follow-up using cost questionnaires. Utility values were based on the EuroQol (EQ-5D-5L), which was administered at baseline, 2, 4, and 6 months follow-up. Utility values were estimated using the Dutch tariff for the EQ-5D-5L[46]. Quality-Adjusted Life-Years (QALYs) were calculated using linear interpolation between measurement points.

The Medical Ethics Review Board of the VU University Medical Centre approved both studies (REALISE-NL35897.029.11 and HypoAware-NL47354.029.13). All participants gave written

informed consent before data collection began. Tables describing baseline characteristics of the REALISE and HypoAware study populations are included in the Appendix (Supplementary Tables 1 and 2). For a detailed description of both studies, the reader is referred elsewhere[42, 45-47].

Statistical analysis

In total, 14 full economic evaluations were performed for both the REALISE and HypoAware study. In the first analysis, a statistical approach was used, in which baseline imbalances, the skewed nature of cost data, the correlation between costs and effects and missing data were ignored. Thus, this approach simply compared the difference in costs and effects between both groups using t-tests, including only participants with complete cost and effect data, while assuming that both costs and effects were normally distributed and that costs and effects were not correlated. Although this statistical approach ignores all of the challenges in trial-based economic evaluations, it is still being used in practice[7, 8, 48]. Step-by-step, the analyses accounted for the different statistical challenges, until in the final approach all of the statistical challenges were accounted for using the following methods:

- Baseline imbalances: Regression-based adjustment was used[16, 49, 50]. Costs and effects were corrected for their baseline value, if available, and for relevant confounding variables. Variables were considered to be a confounder if the estimated regression coefficients for the cost or effect differences changed by 10% or more when the possible confounding factor was added to the model[50, 51]. For the REALISE study, confounders of costs were participants' baseline mental health status, physical health status, risk of future work disability, fear-avoidance beliefs about work, treatment credibility and treatment expectations. Confounders of effects included the participants' baseline utility value, mental health status, back pain, and risk of future work disability. For the HypoAware study, confounders of costs were the participants' baseline costs, number of severe hypoglycemia episodes during the previous 6 months, and wearing a real-time sensor. Confounders of effects comprised the participants' baseline utility value and

marital status.

- Skewed costs: Non-parametric bootstrapping with 5,000 replications was used[33, 52-54]. The non-parametric bootstrap is a data-based resampling method to estimate statistical uncertainty, without making any distributional assumptions[52]. Bootstrapped confidence intervals were estimated using the bias-corrected and accelerated bootstrap method. The advantage of using bias-corrected and accelerated bootstrapping over percentile bootstrapping, is that it adjusts better for skewness and bias of the sampling distribution, resulting in more accurate confidence intervals[52, 55]. In the REALISE study, the skewness of costs was 1.70 and the kurtosis was 5.75 (excess kurtosis 2.75). In the HypoAware study, the skewness of costs was 1.39 and the kurtosis was 3.90 (excess kurtosis 0.90). The positive skewness indicates that the distribution is skewed to the right and the excess kurtosis indicates a relatively long right tail (i.e. relatively many outliers).
- Correlation between costs and effects: Seemingly unrelated regression (SUR) analysis was used in which two separate regression models were specified simultaneously (i.e. one for costs/one for effects)[33, 34]. In the REALISE study, the correlation between costs and effects was = -0.42. In the HypoAware study, the correlation between costs and effects was = -0.44. A negative correlation indicates that individuals with relatively worse outcomes induce higher costs.
- Missing data: Missing data were assumed to be MAR[35]. Multivariate Imputation by Chained Equations (MICE) with predictive mean matching (PMM) was used to predict and impute the missing values based on observed data[26, 56]. PMM was used to deal with the skewed distribution of costs[18]. The advantage of PMM is that it is more robust against non-normal data than linear regression estimation methods, as it uses the observed distribution of the data and non-existing values cannot be imputed[57]. The number of imputed datasets was increased until the loss of efficiency was less than 5%, resulting in 10 imputed datasets for the REALISE study and 20 imputed datasets for the HypoAware study[58]. The imputed datasets were analysed separately

to obtain a set of estimates, which were then pooled using Rubin's rules[35] to obtain overall estimates, variances, and confidence intervals[35, 58, 59]. In the REALISE study, 33 (24%) participants had missing cost data and 21 (15%) had missing effect data. In the HypoAware study, 28 (17%) participants had missing cost data and 20 (12%) had missing effect data.

An overview of the 14 analytical approaches used in this study as well as the statistical challenges they account for can be found in Table 1. For all approaches, incremental costs and QALYs, 95% confidence intervals around incremental costs and QALYs, incremental cost-effectiveness ratios (ICERs) and cost-effectiveness accessibility curves (CEACs) were estimated and compared. ICERs were calculated by dividing incremental mean costs by incremental mean QALYs. CEACs were estimated using the Incremental Net Monetary Benefit (INMB) approach[60]. CEACs represent the probability of an intervention being cost-effective (y-axis) for a range of different ceiling ratios (x-axis) and provide a summary measure of the joint uncertainty surrounding costs and effects[61, 62]. All analyses were performed in StataSE 16® (StataCorp LP, CollegeStation, TX, US).

Comparison of the statistical approaches

Statistical approaches were compared in terms of how sensitive the point estimates are to changes in the statistical approaches (i.e. value sensitivity) and how sensitive the conclusion of an economic evaluation is to changes in statistical approaches (i.e. decision sensitivity)[63]. Value sensitivity was assessed by comparing incremental costs and QALYs, the corresponding confidence intervals, and ICERs across the 14 statistical approaches. Decision sensitivity was assessed by comparing the CEACs of the 14 statistical approaches. For comparing and interpreting the CEACs, thresholds of 0 €/QALY gained, 10,000 €/QALY gained and 23,300 €/QALY gained (i.e. about 20,000 £/QALY gained) were used, which refer to a situation in which decision-makers are not willing to pay anything per QALY gained, the Dutch willingness-to-pay (WTP) thresholds (i.e. between 10,000€/QALY gained and 80,000 €/QALY gained depending on disease severity) and the British National Institute for Health

and Care Excellence (NICE) threshold, respectively.

Table 1. Statistical approaches summarized according to the statistical challenge handled.

Analysis	Baseline imbalances	Skewed costs	Correlated costs and effects	Missing data
1.- Two t-tests (1 for costs; 1 for effects) complete-case analysis	×	×	×	×
2.- Two bootstrapped t-tests; complete-case analysis	×	✓	×	×
3.- Two regressions with correction for confounders; complete-case analysis	✓	×	×	×
4.- Two bootstrapped regressions with correction for confounders; complete-case analysis	✓	✓	×	×
5.- Two t-tests; mean imputation	×	×	×	✓
6.- Two bootstrapped t-tests; mean imputation	×	✓	×	✓
7.- Two regressions with correction for confounders; mean imputation	✓	×	×	✓
8.- Two bootstrapped regressions with correction for confounders; mean imputation	✓	✓	×	✓
9.- Two t-tests; MI	×	×	×	✓
10.- Two bootstrapped t-tests; MI	×	✓	×	✓
11.- Two regressions without correction for confounders; MI	✓	×	×	✓
12.- Two bootstrapped regression with correction for confounders; MI	✓	✓	×	✓
13.- SUR without correction for confounders; MI	×	✓	✓	✓
14.- SUR with correction for confounders; MI	✓	✓	✓	✓

MI Multiple Imputation, SUR Seemingly Unrelated Regression

Legend
×- Not accounted for in the analysis
✓- Accounted for in the analysis

RESULTS

Value sensitivity

In the REALISE study, cost and effect differences ranged from -€782 and -0.001 when ignoring all statistical challenges (analysis 1) to -€82 and 0.011 when accounting for all of them (analysis 14), respectively. The associated ICERs ranged from 636,744 €/QALY gained in analysis 1 to -7,502 €/QALY gained in analysis 14. In analyses 1 to 4, the intervention was less costly and less effective than the control, whereas in analyses 5 to 14 it was less costly and more effective than the control, indicating dominance of the intervention over the control condition. However, in all analyses statistical uncertainty was considerable and all differences in costs and QALYs were statistically non-significant (see Table 2).

In the HypoAware study, cost and effect differences ranged from -€142 and -0.002 when ignoring all statistical challenges (analysis 1) to €462 and 0.010 when accounting for all of them (analysis 14), respectively. The associated ICERs ranged from 90,989 €/QALY gained in analysis 1 to 46,592 €/QALY gained in analysis 14. In analyses 1 to 8, the intervention was less costly and less effective than control, in analyses 5 to 8 the intervention was less costly and more effective than the control, indicating dominance of the intervention over the control condition, whereas in analyses 9 to 14 it was more costly and more effective than the control. However, in all analyses statistical uncertainty was considerable and all differences in costs and QALYs were statistically non-significant (see Table 3).

Decision sensitivity

In the REALISE study, at a willingness-to-pay of 0 €/QALY, 10,000 €/QALY, and 23,300 €/QALY, the probabilities of cost-effectiveness of the intervention as compared to control were 0.67, 0.57, and 0.55, respectively, when ignoring all statistical challenges (analysis 1) and 0.54, 0.57 and 0.59, respectively, when accounting for all of them (analysis 14). In the HypoAware study, at a willingness-to-pay of 0 €/QALY, 10,000 €/QALY, and 23,300€/QALY, the probabilities of cost-effectiveness of the intervention as compared to control were 0.59, 0.53, and 0.51,

respectively, when ignoring all statistical challenges (analysis 1) and 0.27, 0.33, and 0.40, respectively, when accounting for all of them in analysis 14 (Table 4; Figure 1; Figure 2).

Table 2. Results of the cost-effectiveness analyses: REALISE study

Analysis	Societal costs difference (€) (95% CI)	QALY difference (95% CI)	ICER (€/ QALY gained)
1.- Two t-tests (1 for cost; 1 for effects) complete-case analysis	-782 (-4332 to 2768)	-0.001 (-0.070 to 0.0677)	636,744
2.- Two bootstrapped t-tests; complete-case analysis	-782 (-4850 to 2544)	-0.001 (-0.070 to 0.0677)	636,744
3.- Two regressions with correction for confounders; complete-case analysis	-197 (-3692 to 3297)	-0.009 (-0.077 to 0.058)	21,390
4.- Two bootstrapped regressions with correction for confounders; complete-case analysis	-197 (-3643 to 2832)	-0.009 (-0.077 to 0.058)	21,390
5.- Two t-tests; mean imputation	-840 (-3382 to 1701)	0.004 (-0.049 to 0.057)	-231,148
6.- Two bootstrapped t-tests; mean imputation	-840 (-3467 to 1693)	0.004 (-0.049 to 0.056)	-231,148
7.- Two regressions with correction for confounders; mean imputation	-707 (-3197 to 1783)	0.005 (-0.047 to 0.057)	-145,697
8.- Two bootstrapped regressions with correction for confounders; mean imputation	-707 (-3221 to 1636)	0.005 (-0.047 to 0.057)	-231,148
9.- Two t-tests; MI	-141 (-2318 to 2035)	0.009 (-0.053 to 0.071)	-16,425
10.- Two bootstrapped t-tests; MI	-141 (-2398 to 1908)	0.009 (-0.053 to 0.071)	-16,425
11.- Two regressions with correction for confounders; MI	-73 (-2295 to 1073)	0.011 (-0.050 to 0.071)	-6,880
12.- Two bootstrapped regressions with correction for confounders; MI	-73 (-2248 to 1942)	0.011 (-0.050 to 0.071)	-6,880
13.- SUR without correction for confounders; MI	-141 (-849 to 1164)	0.009 (-0.056 to 0.072)	-16,425
14.- SUR with correction for confounders; MI	-82 (-892 to 1586)	0.011 (-0.058 to 0.079)	-7,502

95% CI 95% Confidence Interval ICER Incremental Cost-Effectiveness Ratio, MI Multiple Imputation, SUR Seemingly Unrelated Regression, QALY Quality Adjusted Life-year

Table 3. Results of the cost-effectiveness analyses: HypoAware study

Analysis	Societal costs difference (€) (95% CI)	QALY difference (95% CI)	ICER (€/ QALY gained)
1.- Two t-tests (1 for cost; 1 for effects) complete-case analysis	-142 (-1298 to 1014)	-0.002 (-0.057 to 0.54)	90,989
2.- Two bootstrapped t-tests; complete-case analysis	-142 (-1326 to 986)	-0.002 (-0.057 to 0.54)	90,989
3.- Two regressions with correction for confounders; complete-case analysis	-23 (-1180 to 1134)	-0.009 (-0.041 to 0.03)	2,536
4.- Two bootstrapped regressions with correction for confounders; complete-case analysis	-23 (-1170 to 1096)	-0.009 (-0.041 to 0.03)	2,536
5.- Two t-tests; mean imputation	-105 (-971 to 759)	0.006 (-0.037 to 0.049)	-17,307
6.- Two bootstrapped t-tests; mean imputation	-105 (-1032 to 739)	0.006 (-0.037 to 0.049)	-17,307
7.- Two regressions with correction for confounders; mean imputation	-85 (-954 to 784)	0.009 (-0.020 to 0.038)	-9,208
8.- Two bootstrapped regressions with correction for confounders; mean imputation	-85 (-966 to 770)	0.009 (-0.020 to 0.038)	-9,208
9.- Two t-tests; MI	438 (-1033 to 1910)	0.006 (-0.041 to 0.053)	74,862
10.- Two bootstrapped t-tests; MI	438 (-940 to 1704)	0.006 (-0.041 to 0.053)	74,862
11.- Two regressions with correction for confounders; MI	486 (-995 to 1968)	0.010 (-0.022 to 0.042)	47,305
12.- Two bootstrapped regressions with correction for confounders; MI	486 (-904 to 1747)	0.010 (-0.022 to 0.042)	47,305
13.- SUR without correction for confounders; MI	438 (-936 to 1712)	0.006 (-0.042 to 0.053)	74,862
14.- SUR with correction for confounders; MI	462 (-921 to 1728)	0.010 (-0.022 to 0.0423)	46,592

95% CI 95% Confidence Interval ICER Incremental Cost-Effectiveness Ratio, MI Multiple Imputation, SUR Seemingly Unrelated Regression, QALY Quality Adjusted Life-year

Table 4. Cost–effectiveness probabilities at different willingness to pay thresholds

	REALISE study			HYPOAWARE study		
Analysis	**0 €/QALY**	**10,000 €/QALY**	**23,300 €/QALY**	**0 €/QALY**	**10,000 €/QALY**	**23,300 €/QALY**
1.- Two t-tests (1 for cost; 1 for effects) complete-case	0.67	0.57	0.55	0.59	0.53	0.51
2.- Two bootstrapped t-tests; complete-case	0.66	0.65	0.62	0.59	0.56	0.54
3.- Two regressions with correction for confounders; complete-case	0.54	0.51	0.50	0.52	0.49	0.47
4.- Two bootstrapped regressions with correction for confounders; complete-case	0.54	0.52	0.50	0.52	0.46	0.41
5.- Two t-tests; mean imputation	0.74	0.59	0.57	0.59	0.54	0.54
6.- Two bootstrapped t-tests; mean imputation	0.74	0.73	0.71	0.59	0.62	0.62
7.- Two regressions with correction for confounders; mean imputation	0.71	0.58	0.56	0.58	0.54	0.55
8.- Two bootstrapped regressions with correction for confounders; mean imputation	0.72	0.71	0.69	0.58	0.64	0.68
9.- Two t-tests; MI	0.55	0.53	0.53	0.28	0.42	0.45
10.- Two bootstrapped t-tests; MI	0.55	0.57	0.59	0.26	0.31	0.38
11.- Two regressions with correction for confounders; MI	0.53	0.52	0.53	0.26	0.42	0.46
12.- Two bootstrapped regression with correction for confounders; MI	0.53	0.56	0.58	0.26	0.32	0.40
13.- SUR with correction for confounders; MI	0.58	0.60	0.61	0.28	0.33	0.39
14.- SUR with correction for confounders; MI	0.54	0.57	0.59	0.27	0.33	0.40

MI Multiple Imputation, SUR Seemingly Unrelated Regression

Figure 1. Cost effectiveness accessibility curves indicating the probability of cost-effectiveness at different willingness to pay thresholds in the REALISE study

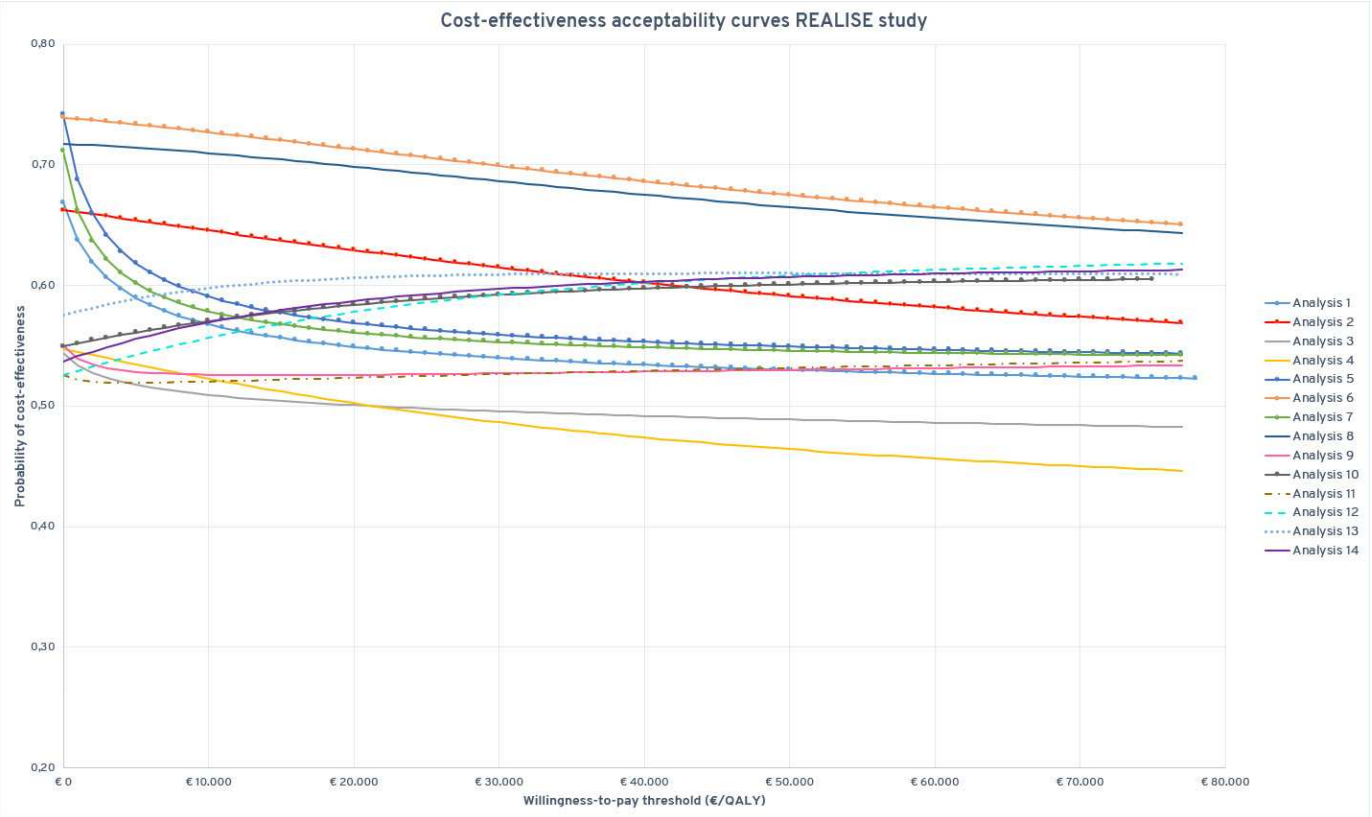
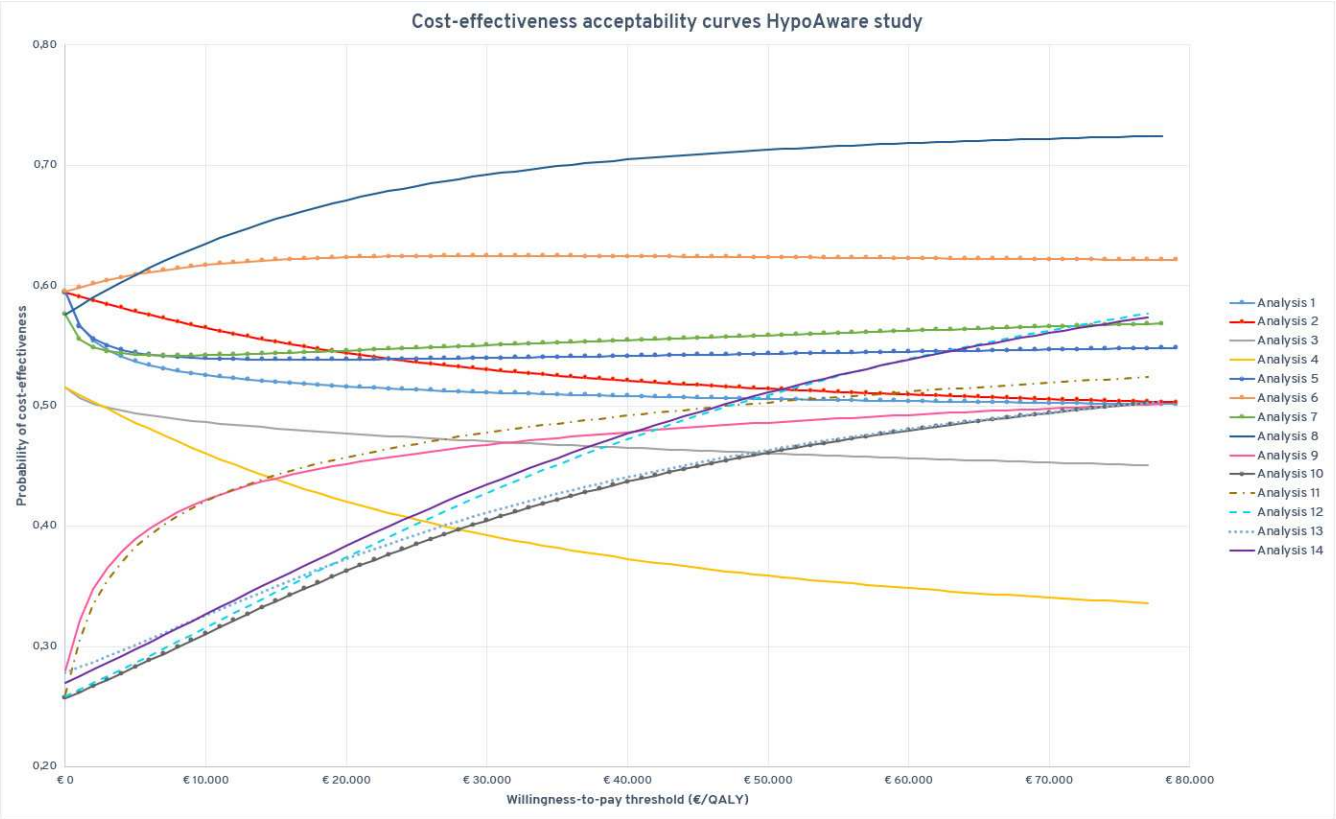


Figure 2. Cost effectiveness accessibility curves indicating the probability of cost-effectiveness at different willingness to pay thresholds in the HypoAware study



DISCUSSION

The current findings indicate that failure to adequately account for baseline imbalances, skewed costs, correlated costs and effects, and missing data in trial-based economic evaluations may have a substantial impact on cost-effectiveness results.

Correction for baseline imbalances showed to have a large impact on the point estimates for both costs and effects, with the impact being most pronounced for costs. When accounting for skewed costs using bootstrapping, in some cases the estimated statistical uncertainty (i.e. confidence intervals) around cost differences increased. However, for the majority of the statistical approaches, the confidence interval width was relatively similar between bootstrapped and non-bootstrapped statistical approaches. Taking into account the correlation between costs and effects had no large effects on the point estimates, nor on the statistical uncertainty surrounding both outcomes. When using different methods to account for missing data, point estimates as well as the amount of statistical uncertainty differed considerably between analyses. These methods consequently had the largest impact on the probabilities of the interventions being cost-effective compared with the control.

Overall, these results indicate that point estimates as well as statistical uncertainty reflected in the probabilities of cost-effectiveness are notably affected when adjusting for baseline imbalances, skewed costs, and/or missing data in trial-based economic evaluations. In the case studies presented, all four statistical challenges were present, indicating that a statistical approach that takes into account all of these challenges simultaneously was the most appropriate approach and is expected to lead to the most valid results and conclusions.

Strengths and limitations

The current study is the first to systematically evaluate the impact of simultaneously adjusting for baseline imbalances, skewed costs, correlated costs and effects, and missing data in trial-based economic evaluations on cost differences, effect differences, ICERs, and statistical uncertainty. Another strength is that, step-by-step, all statistical challenges were accounted for

until all statistical challenges were dealt with, thus also showing the separate impact of accounting for each of these challenges. Finally, all of the applied statistical methods have previously been found to be valid (see references in Box 1). The main limitation of this study is generalizability, as the findings likely depend on the characteristics of the datasets that were analyzed. In addition, the applied statistical methods were employed from a frequentist approach. Bayesian methods are generally more flexible and the interpretation of their results is more intuitive than those of frequentist methods[64-68]. However, Bayesian methods are generally more complex to implement and are less commonly known to most healthcare researchers. Therefore, in line with Gomes et al. [69] we think that frequentist approaches are more likely to improve current practice. Another limitation is that the true outcomes of the REALISE and HypoAware study are not known. Therefore, the performance of the combination of statistical methods used in this study could not be assessed. Finally, although we found a considerable impact of using either one of the statistical approaches on the probabilities of cost-effectiveness in both studies, the conclusions of the studies did not change. Nonetheless, it can be easily imagined that a difference of 0.20 in the probability of cost-effectiveness between analyses 1 and 14 at a ceiling ratio of 10,000 €/QALY (HypoAware study) can lead to a different conclusion in other situations.

Comparison to other studies and implications for further research and practice

One can argue that it is already known that when using different statistical approaches, different results will be obtained. The results of this study reinforce this message and show that it is of utmost importance to align the statistical approach with the statistical challenges identified in a specific dataset. Nonetheless, current practice still shows discrepancies between the statistical approach used and the statistical challenges present in a dataset. For instance, missing data and skewed costs are still inappropriately handled in many trial-based economic evaluations[7, 8, 48], but can have a large impact on the results as illustrated by this study. Amongst others, failure to appropriately handle these issues may be due to a lack of consensus about what

the most optimal methods for dealing with baseline imbalances, skewed costs, correlated costs and effects, and missing data in trial-based economic evaluation are.

Based on the literature, the statistical methods used in this study are currently considered amongst the most appropriate methods. However, statistical methods are in continuous development. For example, multiple imputation is nowadays generally recommended to deal with missing data[19, 35, 70-73]. However, Twisk et al. [74] showed that multiple imputation was not necessary when using longitudinal mixed model analyses to estimate clinical effects, although it is unclear whether this also holds for cost-effectiveness data. Additionally, missing data was assumed to be MAR, however, this assumption might not always hold and data could be MNAR. Recently, an increasing number of guidelines and studies emphasize the importance of checking for possible departure from the MAR assumption[28, 36, 75-77]. It is recommended to perform sensitivity analyses, using other methods such as selection and/or pattern-mixture models[76]. Furthermore, the handling of clustered data or longitudinal data was not investigated in this study, whereas failure to account for clustering will underestimate statistical uncertainty, can lead to inaccurate point estimates, and may in turn lead to incorrect inferences[10, 78]. It is also important to note that the statistical challenges identified in this study are only a selection of possible statistical issues that might arise when analyzing trial-based economic evaluations[79]. Therefore, in order to improve the statistical quality of trial-based economic evaluations, it is helpful to expand the health economic literature by laying out the current state of play regarding statistical methods for trial-based economic evaluations, and to develop guidance and frameworks in which specific statistical methods are recommended to be used in trial-based economic evaluations.

Conclusion

The current study emphasizes the importance of adequately accounting for baseline imbalances, skewed costs, correlated costs and effects, and missing data in trial-based economic evaluations by demonstrating that ignoring them may lead to different cost-effectiveness results. Therefore, when conducting trial-based economic evaluations, it is of utmost importance to

first check the data to identify statistical challenges that need to be accounted for in the analysis, and then adequately deal with them. Furthermore, it is worthwhile to develop consensus among researchers about frameworks and guidelines on how to best analyze trial-based economic evaluations. To facilitate researchers in appropriately dealing with baseline imbalances, skewed costs, correlated costs and effects, and missing data in trial-based economic evaluations, a software code (Stata) is provided in Supplementary File 1 for the most advanced statistical approach.

DISCLOSURES

Code availability Stata syntax is provided as supplementary file 3.

Acknowledgements Both the REALISE and HypoAware studies were funded by the Netherlands Organization for Health Research and Development (REALISE-171102010 and HypoAware - 837001406).

SUPPLEMENTARY FILES

Supplementary table 1. Baseline characteristics REALISE study.

Baseline characteristics	**Control group (SE)** *(n=77)*	**Intervention group (SE)** *(n=92)*
Age in years	46.7 (12.2)	46.9 (11.6)
Female (n, %)	44 (57.1)	54 (58.7)
Living alone (n, %)	8 (10.4)	16 (16.3)
Education (n, %)		
Low	17 (22.1)	20 (21.8)
Middle	35 (45.5)	47 (51.1)
High	25 (32.5)	25 (27.1)
Employment (n, % yes)	57 (74.0)	74 (80.4)
Level of herniation (n, %)		
L2-3	2 (2.6)	1 (1.1)
L3-4	4 (5.2)	10 (10.9)
L4-5	42 (58.3)	31 (33.7)
L5-S1	29 (37.7)	48 (52.2)
L5-6	2 (2.6)	1 (1.1)
Type of herniation (n, %)		
Sequester	34 (44.2)	34 (37.0)
Bulging disc	46 (59.7)	57 (62.0)
Extraforaminal	2 (2.6)	1 (1.1)
Functional status (ODI, 0-100)	50.4 (15.6)	48.6 (17.3)
Pain intensity leg (NRS, 0-10))	7.7 (1.8)	7.8 (1.9)
Pain intensity back (NRS, 0-10))	6.1 (2.6)	6.5 (2.5)
General physical health (SF12, 0-100)	26.7 (15.4)	26.2 (16.1)
General mental health (SF12, 0-100)	50.3 (21.8)	51.6 (21.5)
Psychosocial status (OMPSQ, 0-210)	114.2 (20.5)	109.0 (24.9)
Fear avoidance beliefs physical activity (FABQ, 0-24)	15.4 (5.4)	16.1 (4.4)
Fear avoidance beliefs work (FABQ)	18.5 (11.3)	16.8 (11.0)
Expectation: expectancy surgery (CEQ, 3-27)	22.9 (3.0)	23.2 (2.8)

Continued - Supplementary table 1. Baseline characteristics REALISE study.

Expectations : credibility surgery (CEQ, 3-27))	21.7 (3.7)	22.0 (3.2)
Expectations: credibility item intervention (CEQ 1-9)	6.3 (1.8)	6.5 (1.8)
Expectations: credibility item control (CEQ 1-9)	6.5 (1.4)	6.4 (1.6)
Pain Coping : active (PCI)	6.5 (1.3)	6.7 (1.3)
Pain Coping: passive (PCI)	6.5 (1.2)	6.5 (1.3)
Duration of complaints (n, %)		
0-1months	0	2 (2.2)
1-2 months	3 (3.9)	6 (6.6)
2-3 months	7 (9.1)	1 (1.1)
3-6 months	29 (37.7)	35 (38.0)
6-9 months	13 (16.9)	18 (19.6)
9-12 months	7 (9.1)	6 (6.5)
> 12 months	18 (23.4)	24 (26.1)
Medication use (n, %)		
Every day	47 (61.0)	56 (60.9)
Not every day	14 (18.2)	18 (19.6)
No	16 (20.8)	18 (19.6)
Surgical complications (n, %)		
Nerve root injury	1 (1.3)	1 (1.1)
Dural tear	2 (2.6)	2 (2.2)
Increase in sensimotor deficit	1 (1.3)	0

Supplementary table 2. Baseline characteristics HypoAware study.

Baseline characteristics	**Control group (SE)** *(n=66)*	**Intervention group (SE)** *(n =71)*
Age (years)	51.3 (14.0)	52.7 (12.4)
Female, n (%)	29 (44)	34 (48)
BMI	25.2 (3.9)	26.0 (5.3)
Dutch origin, n (%)	65 (99)	67 (95)
Employed, n (%)	37 (56)	37 (52)
Education, n (%)		
Primary education	16 (24)	23 (32)
Secondary education	30 (46)	20 (28)
Higher education	20 (30)	28 (39)
With partner, n (%)	46 (70)	53 (75)
Type diabetes, n (%)		
Type 1 diabetes,	59 (91)	62 (87)
Type 2 diabetes	6 (9)	8 (11)
Other (MODY)	1 (2)	1 (1)
Treatment, n (%)		
CSII	36 (55)	29 (41)
MDI	30 (46)	42 (59)
Comorbidity, n (%)	33 (50)	40 (56)
HbA1c (mmol/mol)	60.4 (12.2)	60.8 (11.2)
HbA1c (%)	7.7 (1.1)	7.7 (1.0)
Diabetes duration (years)	27.5 (13.1)	24.6 (14.0)
Age of diagnosis (years)	23.9 (12.2)	28.2 (13.9)
Complications, n (%)	28 (42)	29 (41)
Attended a diabetes education program, n (%)	10 (15)	8 (11)
Previous experience with real-time sensor, n (%)	19 (29)	23 (32)
Number of blood glucose measurements per day	4.5 (2.2)	4.6 (2.3)
Non-severe hypoglycemic events per week (,4 mmol/L [,72 mg/dL])*	7.4 (3.9)	5.3 (3.8)

Continued - Supplementary table 2. Baseline characteristics HypoAware study.

Impaired hypoglycemia awareness (Gold score), n (%)	48 (73)	56 (79)
Severe hypoglycemic events in the previous 6 months	1 (0–5)	2 (0–6)

*Data are reported as the mean (SD) or median (IQR), unless otherwise indicated. CSII, continuous subcutaneous insulin infusion; MDI, multiple daily injections. *In participants without RT-CGM at T1 (total n = 98; control n = 45; intervention n = 53).*

Supplementary FILE 3. Syntax with footnote

Syntax:The statistical approach in trial-based economic evaluations matters: get your statistics together!

```
********** CEA ANALYSES **********

clear

set more off

cd "XXX "
capture log close

log using "XXXX .smcl", replace

***** Fill in n = number of imputations *****
local n = XXX

forvalues j=1(1)`n' {
local y = `j'

use "XXXXX `y'", clear
bootstrap bootcost_diff = _b[YYYY :XXXX ] booteffect_diff =
_b[YYYY :XXXX ], reps(XXXX ) seed(XXXX ) saving("boots`y'",
replace) bca: sureg  (YYYY  = XXXX ) (YYYY  = XXXX )

mat betaCE= e(b)        /* extract the matrix of regression
                        coefficients */
mat se = e(se)          /* extract standard errors */
mat limits = e(ci_bc)   /* extract confidence limits */
mat vari = e(V)         /* extract the variance-covariance
                        matrix */

gen cost_diff     = betaCE[1,1]     /* create differential costs */
gen effect_diff   = betaCE[1,2]     /* create differential effects */

gen N             = e(N)            /* extract sample size*/
```

Continued - Supplementary FILE 3. Syntax with footnote

```
gen LL_effect    = limits[1,2]
gen UL_effect    = limits[2,2]
gen LL_cost      = limits[1,1]
gen UL_cost      = limits[2,1]

gen cost_var     = vari[1,1]     /* extract the variance of the
                                    mean differential costs from the
                                    VC matrix */
gen effect_var   = vari[2,2]     /* extract the variance of the
                                    mean differential effect from the
                                    VC matrix */
gen cov          = vari[1,2]     /* extract the covariance
                                    between mean differential costs
                                    and effect */

save postboots`y', replace
}

clear
set more off

cd "XXXX "

***** Fill in n = number of imputations *****
local n = XX

/* append bootstrap samples in 1 file */
use boots1, clear
forvalues k=2(1)`n' {
local z = `k'
append using boots`z'
}
save boots, replace

*** All information from the extra information from bivariate
regression needs to be appended, allowing to pool according to
Rubin's rules ***
```

Continued - Supplementary FILE 3. Syntax with footnote

```
use postboots1, clear
forvalues l=2(1)`n' {
local a = `l'
append using postboots`a'
}

by _mi_m, sort: drop if _n != _N
save postboots, replace

keep cost_diff LL_cost UL_cost cost_var effect_diff LL_effect
UL_effect effect_var cov

append using boots

gen Za=1.95996

/* estimate confidence limits for effects using Rubin's rules */
egen effect_diff_pooled = mean(effect_diff)
egen W=mean(effect_var)
gen _Bdiff=(effect_diff-effect_diff_pooled)^2
egen _Bsum=total(_Bdiff)
gen B=(1/(`n'-1))*_Bsum
gen T=W+(1+(1/`n'))*B
gen seT=sqrt(T)
gen LL_effect_pooled=effect_diff_pooled -(Za*seT)
gen UL_effect_pooled=effect_diff_pooled +(Za*seT)


/* estimate bias-corrected and accelerated confidence limits
for costs */
egen cost_diff_pooled = mean(cost_diff)
egen LL_cost_pooled = mean(LL_cost)
egen UL_cost_pooled = mean(UL_cost)

generate ICER = cost_diff_pooled /effect_diff_pooled

display ICER
display effect_diff_pooled
```

Continued - Supplementary FILE 3. Syntax with footnote

```
display LL_effect_pooled
display UL_effect_pooled
display cost_diff_pooled
display LL_cost_pooled
display UL_cost_pooled

label variable bootcost_diff “Bootstrapped estimates”
label variable cost_diff_pooled “Point estimate”

twoway (scatter bootcost_diff booteffect_diff, msize(small))
(scatter cost_diff_pooled effect_diff_pooled, msize(small)), ///
ytitle(Cost differences (€)) yline(0) xline (0) ///
name(CEplane, replace)

graph save “CEplane.gph”, replace

gen quadrantcompl1 = 0
replace quadrantcompl1 = 1 if bootcost_diff > 0 & booteffect_
diff > 0
replace quadrantcompl1 = 2 if bootcost_diff < 0 & booteffect_
diff > 0
replace quadrantcompl1 = 3 if bootcost_diff < 0 & booteffect_
diff < 0
replace quadrantcompl1 = 4 if bootcost_diff > 0 & booteffect_
diff < 0

label variable quadrantcompl1 “quadrant of CE plane”
label define quadrantcompl1 1 NEQuadrant 2 SEQuadrant 3
SWQuadrant 4 NWQuadrant

sort quadrantcompl1

proportion quadrantcompl1

/* estimate CEA curve using Rubin’s rules */
forvalues i= 0 (1000) 80000 {        /* local macro i counts
                                     from XXXXXX to XXXXXX in
```

Continued - Supplementary FILE 3. Syntax with footnote

```
                                        steps of XXXXXX */
local x = `i'/ 1000                     /* x is created just for
                                        variable names */

gen NB`x'=(`i'*effect_diff)-cost_diff   /* NBs are generated
                                        for each value of i */
gen varNB`x'=`i'^2 * effect_var + cost_var - 2*`i'*cov   /*
                                        variance of NB is
                                        generated */
gen seNB`x'=sqrt(varNB`x')              /* standard error of
                                        NB is generated */

egen meanNB`x'=mean(NB`x')
egen W_NB`x'=mean(varNB`x')
gen _Bdiff_NB`x'=(NB`x'-meanNB`x')^2
egen _Bsum_NB`x'=total(_Bdiff_NB`x')
gen B_NB`x'=(1/(`n'-1))*_Bsum_NB`x'
gen T_NB`x'=W_NB`x'+(1+(1/`n'))*B_NB`x'
gen seT_NB`x'=sqrt(T_NB`x')
local z = meanNB`x'/seT_NB`x'
local prob = normal(`z')

matrix row = (`i',`prob')
matrix ceac = (nullmat(ceac)\row)   /* Matrix containing
                                    probability that
                                    intervention is cost-
                                    effective for each value of i
                                    */
}

svmat ceac              /* The matrix is converted into
                        variables */
matrix drop ceac        /* The unneeded matrix is now
                        dropped */

twoway (line ceac2 ceac1), ytitle(Probability intervention cost-
effective) yscale(range(0 1)) ylabel(0 (0.2) 1) xtitle(Ceiling ratio:
```

Continued - Supplementary FILE 3. Syntax with footnote

€/ QALY) xscale(range(0 80000)) xlabel(0 (10000) 50000)
graph save "CEAC.gph", replace

save postboots, replace

capture log close

1 Fill in path
2 Fill in name log file
3 Fill in number of imputed datasets resulting from multiple imputation procedure
4 Fill in name of imputed dataset
5 Fill in dependent variable = cost
6 Fill in independent variable = treatment arm/grouping variable
7 Fill in dependent variable = effect
8 Fill in independent variable = treatment arm/grouping variable
9 Fill in number of bootstrap replications
10 Fill in seed
11 Note that more variables can be included in the regression equation to adjust for confounders
12 Fill in dependent variable = cost
13 Fill in independent variable = treatment arm/grouping variable
14 Fill in dependent variable = effect
15 Fill in independent variable = treatment arm/grouping variable
16 Fill in path
17 Fill in number of imputed datasets resulting from multiple imputation procedure
18 Change according to desired range of willingness-to-pay values.

REFERENCES

1. Drummond MF, Sculpher, M. J., Claxton, K., Stoddart, G. L., & Torrance, G. W. Methods for the economic evaluation of health care programmes: Oxford university press; 2015.

2. Barbieri M, Drummond, M., Rutten, F., Cook, J., Glick, H. A., Lis, J., Reed, S.D., Sculpher, M., Severens, J.L., & ISPOR Good Research Practices Economic Data Transferability Task Force. What do international pharmacoeconomic guidelines say about economic data transferability? Value in Health. 2010;13(8):1028-37.

3. van Lier LI, Bosmans, J. E., van Hout, H. P., Mokkink, L. B., van Den Hout, W. B., de Wit, G. A., Dirksen, C.D., Nies, H.L.G.R., Hertogh, C.M.P.M., & van der Roest, H. G. Consensus-based cross-European recommendations for the identification, measurement and valuation of costs in health economic evaluations: a European Delphi study. The European Journal of Health Economics. 2018:1-16.

4. Ramsey SD, Willke, R. J., Glick, H., Reed, S. D., Augustovski, F., Jonsson, B., Briggs A., & Sullivan, S. D. Cost-effectiveness analysis alongside clinical trials II—an ISPOR Good Research Practices Task Force report. Value in Health. 2015;18(2):161-72.

5. Adetugbo K, & Williams, H. How well are randomized controlled trials reported in the dermatology literature? Archives of dermatology. 2000;136(3):381-5.

6. Montané E, Vallano, A., Vidal, X., Aguilera, C., & Laporte, J. R. Reporting randomised clinical trials of analgesics after traumatic or orthopaedic surgery is inadequate: a systematic review. BMC clinical pharmacology. 2010;10(1):2.

7. El Alili M, van Dongen, J. M., Huirne, J. A., van Tulder, M. W., & Bosmans, J. E. Reporting and analysis of trial-based cost-effectiveness evaluations in obstetrics and gynaecology. Pharmacoeconomics. 2017;35(10):1007-33.

8. Doshi JA, Glick, H. A., & Polsky, D. Analyses of cost data in economic evaluations conducted alongside randomized controlled trials. Value in Health. 2006;9(5):334-40.

9. Gabrio A, Mason, A. J., & Baio, G. Handling missing data in within-trial cost-effectiveness analysis: a review with future recommendations. PharmacoEconomics-open. 2017;1(12):79-97.

10. Gomes M, Grieve, R., Nixon, R., & Edmunds, W. J. Statistical methods for cost-effectiveness analyses that use data from cluster randomized trials: a systematic review and checklist for critical appraisal. Medical Decision Making. 2012;32(1):209-20.

11. Díaz-Ordaz K, Kenward, M. G., Cohen, A., Coleman, C. L., & Eldridge, S. Are missing data adequately handled in cluster randomised trials? A systematic review and guidelines. Clinical Trials. 2014;11(5):590-600.

12. Ramsey S, Willke, R., Briggs, A., Brown, R., Buxton, M., Chawla, A., Cook, J., Glick, H., Liljas, B., Petitti, D., & Reed, S. Good research practices for cost-effectiveness analysis alongside clinical trials: the ISPOR RCT-CEA task force report. Value in health. 2005;8(5):521-33.

13. Faria R, Gomes, M., Epstein, D., & White, I. R. A guide to handling missing data in cost-effectiveness analysis conducted within randomised controlled trials. Pharmacoeconomics,. 2014;32(12):1157-70.

14. Hoch JS, Briggs, A. H., & Willan, A. R. Something old, something new, something borrowed, something blue: a framework for the marriage of health econometrics and cost effectiveness analysis. Health economics. 2002;11(5):415-30.

15. Thompson SG, & Nixon, R. M. How sensitive are cost-effectiveness analyses to choice of parametric distributions? Medical Decision Making,. 2005;25(4):416-23.

16. van Asselt AD, van Mastrigt, G. A., Dirksen, C. D., Arntz, A.,

Severens, J. L., & Kessels, A. G. How to deal with cost differences at baseline. Pharmacoeconomics. 2009;27(6):519-28.

17. Vroomen, JM, Eekhout, I., Dijkgraaf, M. G., van Hout, H., de Rooij, S. E., Heymans, M. W., & Bosmans, J. E. Multiple imputation strategies for zero-inflated cost data in economic evaluations: which method works best? The European Journal of Health Economics. 2016;17(8):939-50.

18. White IR, Royston, P., & Wood, A. M. Multiple imputation using chained equations: issues and guidance for practice. Statistics in medicine. 2011;30(4):377-99.

19. Dongen JMv, El Alili M, Varga AN, Guevara Morel AE, Jornada Ben A, & Khorrami M, et al. What do national pharmacoeconomic guidelines recommend regarding the statistical analysis of trial-based economic evaluations? Expert review of pharmacoeconomics & outcomes research. E-pub, 2019.

20.Manca A, Hawkins, N., & Sculpher, M. J. Estimating mean QALYs in trial-based cost-effectiveness analysis: the importance of controlling for baseline utility. Health economics. 2005;14(5): 487-96.

21. Nixon RM, & Thompson, S. G. Methods for incorporating covariate adjustment, subgroup analysis and between-centre differences into cost-effectiveness evaluations. Health economics. 2005;14(12):1217-29.

22.Sekhon JS, & Grieve, R. D. A matching method for improving covariate balance in cost-effectiveness analyses. Health economics. 2012;21(6):695-714.

23.Manca A, Hawkins N, & Sculpher MJ. Estimating mean QALYs in trial-based cost-effectiveness analysis: the importance of controlling for baseline utility. Health economics. 2005;14(5):487-96.

24.Nixon RM, Wonderling D, & Grieve RD. Non-parametric methods for cost-effectiveness analysis: the central limit theorem

and the bootstrap compared. Health Econ. 2010;19(3):316-33.

25. van Dongen JM, van Wier, M. F., Tompa, E., Bongers, P. M., van der Beek, A. J., van Tulder, M. W., & Bosmans, J. E. Trial-based economic evaluations in occupational health: principles, methods, and recommendations. Journal of occupational and environmental medicine. 2014;56(6):563.

26.Petrou S, & Gray, A. Economic evaluation alongside randomised controlled trials: design, conduct, analysis, and reporting. Bmj. 2011;342(1548).

27. Baio G. Bayesian Models for Cost-Effectiveness Analysis in the Presence of Structural Zero Costs. Value Health. 2014;17(7):A584.

28.Gabrio A, Mason AJ, & Baio G. A full Bayesian model to handle structural ones and missingness in economic evaluations from individual-level data. Stat Med. 2019;38(8):1399-420.

29.Ng ES-W, Grieve R, & Carpenter JR. Two-stage nonparametric bootstrap sampling with shrinkage correction for clustered data. The Stata Journal. 2013;13(1):141-64.

30.Polsky D, Glick, H. A., Willke, R., & Schulman, K. Confidence intervals for cost–effectiveness ratios: a comparison of four methods. Health economics. 1997;6(3): 243-52.

31. Willan AR, Briggs, A. H., & Hoch, J. S. Regression methods for covariate adjustment and subgroup analysis for non censored cost-effectiveness data. Health economics. 2004;13(5):461-75.

32.Glick H, & Doshi, J. Designing economic evaluations in clinical trials. Economic Evaluation in Clinical Trials. Oxford, UK: Oxford University Press; 2010.

33.Willan AR. Statistical analysis of cost-effectiveness data from randomized clinical trials. Expert review of pharmacoeconomics & outcomes research. 2006;6(3):337-46.

34.Willan AR, Briggs AH, & Hoch JS. Regression methods for covariate adjustment and subgroup analysis for non-censored cost-effectiveness data. Health Econ. 2004;13(5):461-75.

35.Rubin DB. Multiple imputation for nonresponse in surveys. New York: Wiley; 1987. xxix, 258 p. p.

36.Leurent B, Gomes M, & Carpenter JR. Missing data in trial-based cost-effectiveness analysis: An incomplete journey. Health Econ. 2018;27(6):1024-40.

37. Eekhout I, de Vet HC, Twisk JW, Brand JP, de Boer MR, & Heymans MW. Missing data in a multi-item instrument were best handled by multiple imputation at the item score level. J Clin Epidemiol. 2014;67(3):335-42.

38.White IR, & Carlin JB. Bias and efficiency of multiple imputation compared with complete-case analysis for missing covariate values. Stat Med. 2010;29(28):2920-31.

39. Lipsitz SR, Fitzmaurice, G. M., Ibrahim, J. G., Sinha, D., Parzen, M., & Lipshultz, S. Joint generalized estimating equations for multivariate longitudinal binary outcomes with missing data: an application to acquired immune deficiency syndrome data. Journal of the Royal Statistical Society: Series A (Statistics in Society). 2009;172(1): 3-20.

40.Noble SM HW, & Tilling K. Missing data in trial-based cost-effectiveness analysis: the current state of play. Health economics. 2012;21:187-200.

41. Zhang Z. Missing data imputation: focusing on single imputation. Annals of translational medicine. 2016;4(9).

42.Oosterhuis T, van Tulder, M., Peul, W., Bosmans, J., Vleggeert-Lankamp, C., Smakman, L., Arts, M. & Ostelo, R. Effectiveness and cost-effectiveness of rehabilitation after lumbar disc surgery (REALISE): design of a randomised controlled trial. BMC musculoskeletal disorders. 2013;14(1):124.

43.Hakkaart-van Roijen L, Tan S, & Bouwmans C. Handleiding voor kostenonderzoek, methoden en standaard kostprijzen voor economische evaluaties in de gezondheidszorg. Geactualiseerde versie. 2010;2010.

44.Lamers L, Stalmeier P, McDonnell J, & Krabbe P. Measuring the quality of life in economic evaluations: the Dutch EQ-5D tariff. Nederlands tijdschrift voor geneeskunde. 2005;149(28):1574-8.

45.de Wit M, Rondags, S. M. P. A., van Tulder, M. W., Snoek, F. J., & Bosmans, J. E. Cost-effectiveness of the psycho-educational blended (group and online) intervention HypoAware compared with usual care for people with Type 1 and insulin-treated Type 2 diabetes with problematic hypoglycaemia: analyses of a cluster-randomized controlled trial. Diabetic Medicine. 2018;35(2):214-22.

46.Rondags SM, de Wit, M., van Tulder, M. W., Diamant, M., & Snoek, F. J. HypoAware-a brief and partly web-based psycho-educational group intervention for adults with type 1 and insulin-treated type 2 diabetes and problematic hypoglycaemia: design of a cost-effectiveness randomised controlled trial. BMC endocrine disorders. 2015;15(43).

47. Rondags SM, de Wit, M., Twisk, J. W., & Snoek, F. J. Effectiveness of HypoAware, a brief partly web-based psychoeducational intervention for adults with type 1 and insulin-treated type 2 diabetes and problematic hypoglycemia: a cluster randomized controlled trial. Diabetes Care. 2016;39(12):2190-6.

48.Gomes M, Grieve R, Nixon R, & Edmunds WJ. Statistical methods for cost-effectiveness analyses that use data from cluster randomized trials: a systematic review and checklist for critical appraisal. Med Decis Making. 2012;32(1):209-20.

49. Hoch JS, Briggs AH, & Willan AR. Something old, something new, something borrowed, something blue: a framework for the marriage of health econometrics and cost-effectiveness analysis. Health Econ. 2002;11(5):415-30.

50. McNeish DM. Analyzing clustered data with OLS regression: The effect of a hierarchical data structure. Multiple Linear Regression Viewpoints. 2014;40(1):11-6.

51. Twisk JW. Inleiding in de toegepaste biostatistiek. Houten: Bohn Stafleu van Loghum; 2017.

52. Barber JA, & Thompson SG. Analysis of cost data in randomized trials: an application of the non-parametric bootstrap. Stat Med. 2000;19(23):3219-36.

53. Doshi JA, Glick HA, & Polsky D. Analyses of Cost Data in Economic Evaluations Conducted Alongside Randomized Controlled Trials. Value in Health. 2006;9(5):334-40.

54. Ramsey SD, Willke RJ, Glick H, Reed SD, Augustovski F, Jonsson B, et al. Cost-effectiveness analysis alongside clinical trials II-An ISPOR Good Research Practices Task Force report. Value Health. 2015;18(2):161-72.

55. Briggs AH, Wonderling DE, & Mooney CZ. Pulling cost-effectiveness analysis up by its bootstraps: a non-parametric approach to confidence interval estimation. Health Econ. 1997;6(4):327-40.

56. Glick H, Doshi JA, Sonnad SS, & Polsky D. Economic evaluation in clinical trials. Second edition. ed. Oxford: Oxford University Press; 2015. x, 252 pages p.

57. Vink G, Frank, L. E., Pannekoek, J., & Van Buuren, S. Predictive mean matching imputation of semicontinuous variables. Statistica Neerlandica. 2014;68(1):61-90.

58. Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ. 2009;338:b2393.

59. White IR, Royston P, & Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. Stat

Med. 2011;30(4):377-99.

60. Hoch JS DC. Advantages of the net benefit regression framework for economic evaluations of interventions in the workplace: a case study of the cost-effectiveness of a collaborative mental health care program for people receiving short-term disability benefits for psychiatric disorders. J Occup Environ Med. 2014;56(4):441-5.

61. Fenwick E, O'Brien BJ, & Briggs A. Cost-effectiveness acceptability curves--facts, fallacies and frequently asked questions. Health Econ. 2004;13(5):405-15.

62.van Hout BA, Al MJ, Gordon GS, & Rutten FF. Costs, effects and C/E-ratios alongside a clinical trial. Health Econ. 1994;3(5):309-19.

63.Griffin SC. Dealing with uncertainty in the economic evaluation of health care technologies. York, United Kingdom: University of York; 2010.

64.Al MJ, & Van Hout BA. A Bayesian approach to economic analyses of clinical trials: the case of stenting versus balloon angioplasty. Health Econ. 2000;9(7):599-609.

65.Gabrio A, Baio G, & Manca A. Bayesian Statistical Economic Evaluation Methods for Health Technology Assessment. Oxford Research Encyclopedia of Economics and Finance2019.

66.Gabrio A, Hunter R, Mason AJ, & Baio G. Pitfalls of adjusting for mean baseline utilities/costs in trial-based cost-effectiveness analysis with missing data. arXiv preprint arXiv:180507149. 2018.

67. Grieve R, Nixon R, & Thompson SG. Bayesian hierarchical models for cost-effectiveness analyses that use data from cluster randomized trials. Medical Decision Making. 2010;30(2):163-75.

68.Mason AJ, Gomes M, Grieve R, & Carpenter JR. A Bayesian framework for health economic evaluation in studies with missing data. Health economics. 2018;27(11):1670-83.

69. Gomes M, Ng ES, Grieve R, Nixon R, Carpenter J, & Thompson SG. Developing appropriate methods for cost-effectiveness analysis of cluster randomized trials. Med Decis Making. 2012;32(2):350-61.

70. Burton A, Billingham LJ, & Bryan S. Cost-effectiveness in clinical trials: using multiple imputation to deal with incomplete cost data. Clin Trials. 2007;4(2):154-61.

71. El Alili M, van Dongen JM, Huirne JAF, van Tulder MW, & Bosmans JE. Reporting and Analysis of Trial-Based Cost-Effectiveness Evaluations in Obstetrics and Gynaecology. PharmacoEconomics. 2017;35(10):1007-33.

72. Faria R, Gomes M, Epstein D, & White IR. A Guide to Handling Missing Data in Cost-Effectiveness Analysis Conducted Within Randomised Controlled Trials. PharmacoEconomics. 2014;32(12):1157-70.

73. White IR, Royston P, & Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. Statistics in Medicine. 2011;30(4):377-99.

74. Twisk J, de Boer M, de Vente W, & Heymans M. Multiple imputation of missing values was not necessary before performing a longitudinal mixed-model analysis. J Clin Epidemiol. 2013;66(9):1022-8.

75. Faria R, Gomes M, Epstein D, & White IR. A guide to handling missing data in cost-effectiveness analysis conducted within randomised controlled trials. Pharmacoeconomics. 2014;32(12):1157-70.

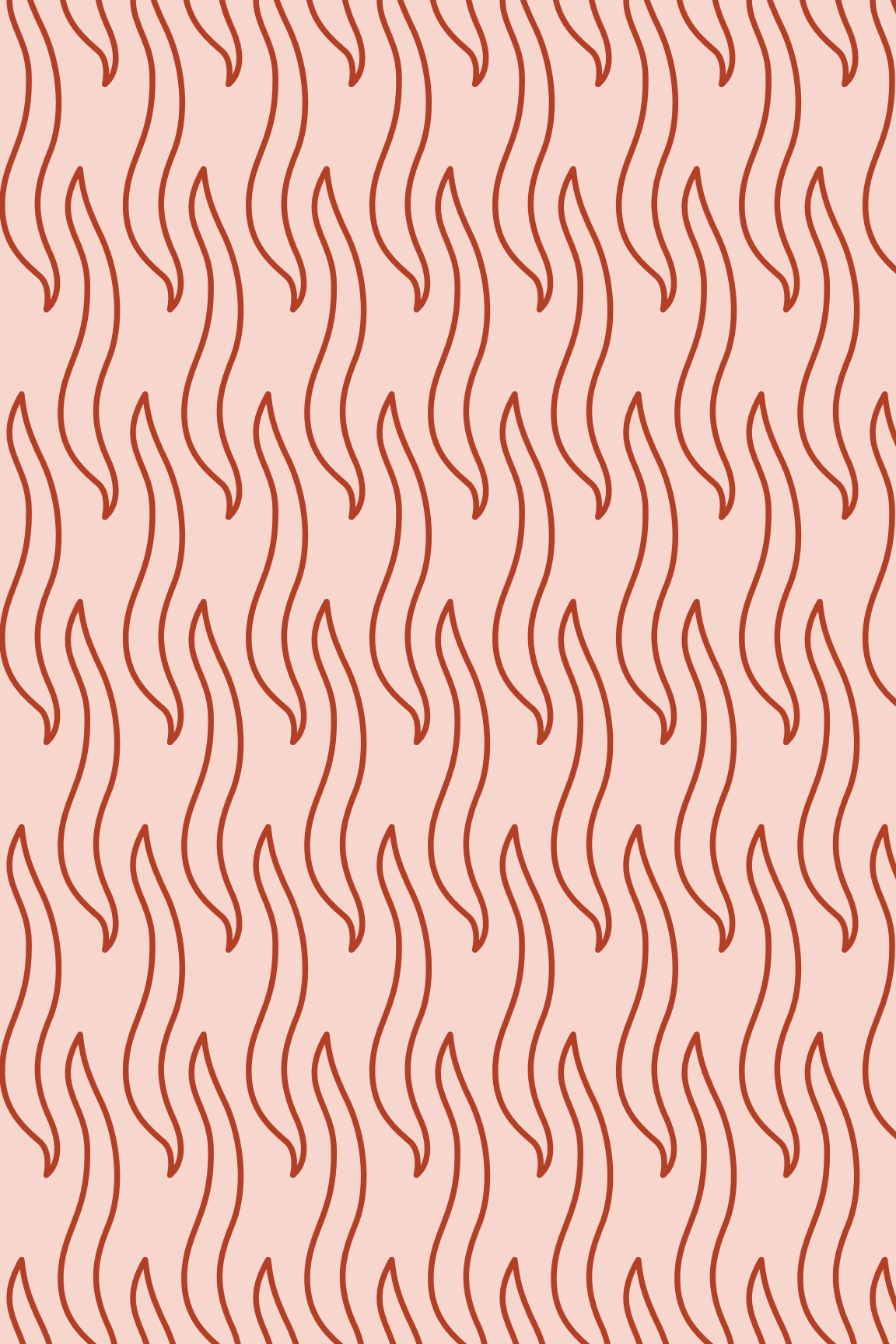
76. Leurent B, Gomes M, Cro S, Wiles N, & Carpenter JR. Reference-based multiple imputation for missing data sensitivity analyses in trial-based cost-effectiveness analysis. Health Econ. 2020;29(2):171-84.

77. Leurent B, Gomes M, Faria R, Morris S, Grieve R, &

Carpenter JR. Sensitivity Analysis for Not-at-Random Missing Data in Trial-Based Cost-Effectiveness Analysis: A Tutorial. Pharmacoeconomics. 2018;36(8):889-901.

78. Grieve R, Nixon, R., & Thompson, S. G. Bayesian hierarchical models for cost-effectiveness analyses that use data from cluster randomized trials. Medical Decision Making. 2010;30(2):163-75.

79. van Dongen JM, El Alili M, Varga AN, Guevara Morel AE, Jornada Ben A, & Khorrami M, et al. What do national pharmacoeconomic guidelines recommend regarding the statistical analysis of trial-based economic evaluations? Expert review of pharmacoeconomics & outcomes research. 2020;20(1):27-37.



General Discussion & Summary

9

General Discussion

GENERAL DISCUSSION

Low back pain is a very widespread health complaint, and is a burden to patients and society. In the Netherlands, total societal costs from low back pain were estimated to be as high as 3.5 billion euros in 2007, which equals about 0.6% of the Dutch gross national product[1]. Other countries have also reported high annual societal costs due to low back pain, namely; the United States, more than 100 billion dollars [2,3], 9.0 billion Australian Dollars in Australia [4], 6.6 billion Euros in Switzerland [5], and 12.3 billion British Pounds in the UK [6,7]. These costs are driven mostly by work absenteeism and a loss of productivity while being at work [8]. Low back pain patients with higher pain levels were found to be more absent from work [9]. When low back pain persists for a longer period, it may result in higher functional disability, higher absenteeism, poor health-related quality of life and high cost.

Therefore, the overall aim of the thesis was to contribute to the development of a sound evidence base on:

1) the relationship between low back pain, outcomes, and costs *(Chapter 2 and 3)*.
2) the effectiveness and cost-effectiveness of sciatica treatments *(Chapter 4, 5 and 6)*.

And to improve:

3) scientific methods in low back pain research *(Chapter 7 and 8)*.

In this general discussion, the results of this thesis will be summarized and discussed. This will be done in three separate, but related sections.

The first section *(Theme A)* is entitled *Relationship between low back pain, outcomes, and costs*. In this theme, the potential interaction of pain and disability is explored as well as the potential combined influence of pain and disability on costs and health-related quality of life. Additionally, a selection of methodological considerations with regard to the generalizability and the analytic strategy of *Chapter 2* and *Chapter 3* are further explored.

The second section *(Theme B)* is entitled *Effectiveness and cost-effectiveness of sciatica treatments*. In this theme, the

effectiveness and cost-effectiveness of combination therapy, consisting of MDT and TESIS, compared to no intervention while being on the waiting list for lumbar herniated disc surgery (i.e. usual care), is explored *(Chapter 4 and 5)*. Additionally, the challenges associated with the low patient inclusion rate in our randomized control trial (i.e. the PLUS-study) are discussed from an organization, neurosurgeon, research design, and patient point of view. Further, the generalizability of the findings of the effectiveness and cost-effectiveness analysis is explored. The effectiveness of exercises on sciatica is also explored *(Chapter 6)*.

The third section *(Theme C)* is entitled *Methodological studies*. In this theme, the construct validity of the Global Perceived Effect (GPE) scale for measuring self-perceived recovery in patients with sciatica is explored *(Chapter 7)*. Additionally, it is discussed whether the GPE is a true transition scale or if the score on the GPE is influenced by current health status. *Chapter 8* evaluated whether accounting for statistical challenges such as; missing data, skewed costs, the correlation between costs and effects, and baseline imbalances in trial-based economic evaluations has an impact on cost-effectiveness results.

THEME A: RELATIONSHIP BETWEEN LOW BACK PAIN, OUTCOMES AND COSTS

Research questions:

1. What is the association between pain severity/disability with health-related quality of life and costs within and between individuals over a 3-month period? *(Chapter 2)*
2. Which factors predict high societal costs among chronic low back pain patients? *(Chapter 3)*

Main findings

Chapter 2 assessed the longitudinal association between pain severity and disability with health-related quality of life and costs. The results showed that pain severity and disability were both longitudinally related to health-related quality of life, societal costs, and healthcare costs. To illustrate, a clinically relevant improvement in disability of 10 points on the ODI [10] was associated with an improvement in health-related quality of life by 0.096 points (range 0–1) and a decrease in societal costs and healthcare costs by €170 and €80 per 3-month period, respectively. The decrease in health-related quality of life is bigger than the minimal clinically important difference for this outcome (i.e. 0.057) [11,12], whereas the decreases in societal costs and healthcare costs highlight big potential savings. It is noteworthy that disability compared to pain had a higher impact on health-related quality of life and costs, even though patients had high baseline scores of pain (i.e. [Mean: 73; SD:16] range 0-100) and relatively 'lower scores' on disability (i.e. [ODI mean:11; SD:9] range 0-100) [13].

Chapter 3 assessed which factors predict high societal costs among chronic low back pain patients. High impact of pain experience, being female, non-Dutch nationality, a combined diagnosis (low back pain caused by both facet joints and intervertebral disc), poor physical health, high functional disability, low health-related quality of life, younger age, and decreasing pain intensity were factors found to increase the odds of having high societal costs. The model's overall fit was good, whereas its explained variance was relatively low [14-16]. That is,

only 14.3% of the variance in high societal costs was explained by the identified predictive factors. Two sensitivity analyses were performed using different cut-off points for high societal costs. Poor physical health, high functional disability, low health-related quality of life, high impact of pain experience, non-Dutch nationality and decreasing pain were found to be predictive of having high societal costs in all models, making them the most robust predictors of high societal costs among chronic low back pain patients.

Discussion

Methodological considerations

Methodological strengths and limitations of the longitudinal study *(Chapter 2)* and prediction model *(Chapter 3)*, including their internal and external validity, have been discussed in both chapters. Nevertheless, a selection of methodological considerations with regard to generalizability, the potential interaction between pain and disability, and the analytic strategy require further exploration.

Generalizability of results

Both *Chapter 2* and *Chapter 3* were based on the same dataset, i.e. that of the observational part of the MINT study [17]. Patients included in the observational study did not want to or were not eligible to participate in the randomised control trials of the MINT study. The patients were aged between 18 and 70 years, were referred to a pain clinic with suspected chronic mechanical low back pain and did not experience an improvement of symptoms after conservative treatment [17]. Patients who had, for example, a body mass index higher than 35 were excluded from the randomized clinical trials of the MINT study and were included in the observational study [17]. This resulted in a relatively high number of overweight patients. To illustrate, of the 6,316 chronic low back pain patients in the observational study group, 67% were overweight, which is higher than the 58% of chronic low back patients being overweight in the study of van Dongen et.al (2017). These numbers, however might also not be completely representative of all chronic low back pain patients in the Netherlands, because of the overrepresentation of for example

overweight in the observational group. Nonetheless, the study by van Dongen et.al (2017), at least assessed all chronic low back pain patients visiting an orthopedic clinic in the Netherlands, instead of just a selection of those patients [18]. Other factors that can limit generalizability of the findings to other healthcare settings and other types of low back pain include the over-representation of females in our study compared to the general low back pain population (i.e. 66% versus 56-58%), the under-representation of low educated patients (i.e. 56% versus 71%) [13,18,19], and the secondary setting of the current study. However, we assume that our findings are generalizable to other secondary healthcare settings with a similar low back pain population and with similar health care issues. In addition, the MINT study was a multicenter trial conducted at 16 multidisciplinary pain clinics in the Netherlands. Enrolling patients from a large number of different centers increases generalizability of results to similar patient populations [20,21].

In *Chapter 3*, only four (2.8%) out of the 171 patients in the high-cost group in the main analysis were non-Dutch nationals. It is unlikely that these four participants are representative of all non-Dutch nationals with low back pain in the Netherlands. Hence, even though non-Dutch nationality was identified as a predictor in all of the models, further research is needed to establish whether non-Dutch nationality is indeed a very strong predictor of having high societal costs among low back pain patients. Nevertheless, migrants compared to the majority of the people in the host country usually have poorer health outcomes [22]. To illustrate, immigrants were previously found to be more likely to go through the whole rehabilitation program and not get better or seek care in a later stadium thereby attracting more costs [22]. This indicates that health status and health behavior differs between immigrants and nationals of a given country [23]. These differences also exist among immigrants themselves, and can be partly explained by social economic and demographic variables [22]. Of the four non-Dutch nationals in the high costs group, three had a low level of education and one had a high level of education. Low education level or economic status were not predictors of high societal costs in our study. This indicates that there are possibly more factors, such as barriers immigrants encounter in the health system, being over-represented in

manual labor jobs, health insurance, and type of employment accessible to non-Dutch nationals. Of them, health insurance was not a predictor of high societal costs in our study *(Chapter 3)*. This could be because the basic health insurance package is compulsory for all in the Netherlands. However, differences in health insurance might be a factor in other countries, particularly those without public and mandated health insurance. Therefore, it is important to further research on the differences among non-Dutch nationals who made high societal costs in *Chapter 3*.

The relationship between race/nationality/ethnicity and low back pain symptoms and outcomes is complex [24-26]. There are differences in how racial and ethnic minorities cope with pain-related conditions compared to the racial/ethical majority [24,25], hence affecting their healthcare seeking behavior, and consequently the course of chronic pain and costs. Ethnicity is probably influential in how individuals shape their pain experiences, because pain is shaped by the interactions among biological, psychological, and social variables [25]. Also, pain severity, behavioral, and emotional responses to pain may differ with ethnicity [25]. In the United States, ethnic differences in low back pain are noted [25]. Carey and Garette (2003) reported that black people compared to white people in North Carolina in the United States reported worse functional disability, but received less intense treatments (i.e. radiographs, advances imagining) even after income, insurance status, education, and baseline severity of low back pain was controlled for [26]. The authors reported that healthcare providers may perceive the presentation of black people and white people with musculoskeletal problems quite differently. Racial discrimination in pain management could result from, among others, differences in pain expression and pain interpretation, language barriers, cultural barriers, and racism [27]. In light of the recent discussions about diversity and racism, it would be good if grant agencies prioritize this topic. Further research could aim to understand if nationality is really a strong predictor of high costs and explore effective strategies to reduce high costs.

It is also noteworthy that some of the results in *chapter 3* seem counterintuitive, particularly the finding that lower pain levels were predictive for higher costs. A possible explanation for this might be that only chronic low back pain patients in a

secondary setting were included in the MINT study, whereas patients in primary setting are reported to have higher pain levels than those in secondary settings [28]. This is due to the fact that patients seen in primary settings are often acute and have higher pain levels. In addition, patients in secondary settings usually receive more expensive and/or extensive treatment, i.e. surgery, which tends to offer quicker relief to their complaints. As a consequence, it is unknown whether our finding that lower pain levels were predictive for higher costs is also applicable to acute and sub-acute low back pain patients with low levels of pain.

Combined influence of disability and pain on costs and health-related quality of life

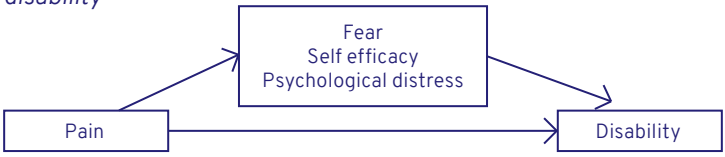
In *chapter 2*, only the separate *longitudinal* relationships of pain and disability with healthcare costs, societal costs, and health-related quality of life were explored. This was done because only cross-sectional evidence on these relationships was available at the moment. As a consequence, the relationship between disability and pain was not explored and neither was the possible mediating effect of disability in the relationship between pain and health-related quality of life and costs. A mediator is defined as a variable that affects the relationship or association between variables (i.e. independent and outcome variables)*(Figure 1)*.

Figure 1: The conceptual model showing a mediator effect [29]

Pain is generally thought to precede disability, therefore a better understanding of the development of disability in low back pain patients can be reached when causal pathways between pain and disability are better understood. In the dataset used in *chapter 2* and *chapter 3*, the reported levels of disability were lower than those of pain. Nonetheless, disability was found to be more highly correlated with costs and health-related quality of life compared to pain, further highlighting the importance of understanding the relationship between these two. Various models have been proposed to explain how pain leads to disability. The “fear of

avoidance model of pain" suggests that pain-related fear triggers mechanisms that result in movement and activity avoidance, which in the long-term can cause disability [30]. Another potential explanation is the "social cognitive theory" which describes the relationship between self-efficacy, expectancies, intention, and behaviour. Pain-related self-efficacy is defined as the beliefs that chronic pain patients have regarding their ability to carry out activities despite their pain [31]. However, there is a lack of high-quality evidence on recognised mediators, which may explain how pain leads to disability [32]. A systematic review by Lee et al concluded that self-efficacy, psychological distress, and fear mediate the relationship between pain and disability in people with low back pain *(Figure 2)*. However, no definitive conclusions regarding causality on the development of disability could be made due to poor quality of the included studies and the limited number of studies. Therefore, future studies should investigate the causal pathways between disability and pain in low back pain.

Figure 2: Possible mediating relationship between pain and disability



Another possible mediating effect worth investigating is that of disability in the relationship between pain and health-related quality of life and costs *(Figure 3)*. Literature suggests that pain is related to disability and in our study we found that disability is related to costs and health-related quality of life [13]. This might suggest that disability mediates the relationship between pain and costs and health-related quality of life. However, further research is warranted to explore the possible mediating effect.

Figure 3: Possible mediating relationship between pain and costs/ Health related quality of life.

Analytical strategy

In *Chapter 2*, we opted for a Generalized Estimating Equation (GEE) analysis to assess the longitudinal relationship between pain severity and disability with health-related quality of life and costs, because we assumed that its unit-specific correlation structure would be most realistic. As a result, however, our GEE estimates highly depend on the assumptions made prior to the analyses. Namely, that with a GEE analysis, the adjustment for time is carried out by assuming *a priori* a certain "working" correlation structure for the repeated measurements. Even though GEE analysis is presumed to be robust against a wrong choice of correlation structure, evidence suggests that results may differ extensively across correlation structures [33]. In *Chapter 2* we assumed an "exchangeable" correlation structure *(Figure 4)*, in which correlations between subsequent measurements are equal irrespective of the length of the time intervals [33]. Our assumptions were based on Twisk et al (1997) who recommended using the simplest correlation structure that fits the data well [33]. The robustness of our findings regarding the choice of correlation structure, were assessed by means of a post-hoc analysis with an "unstructured" correlation structure *(Figure 5)*. With an "unstructured" correlation structure no particular structure is assumed and all possible correlations between repeated measurements have to be estimated [33]. As the results of the post-hoc analyses were in line with those of the main analysis we considered our findings to be robust.

Figure 4: Exchangeable correlation structure (any two responses have the same correlation (p), only one correlation parameter to be estimated, order of observation within the cluster is arbitrary)

$$\begin{bmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{bmatrix}$$

Figure 5: Unstructured correlation structure (observations are correlated with no assumption of structure)

$$\begin{bmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{bmatrix}$$

In *Chapter 3*, we opted for logistic regression instead of machine learning for developing the prediction model. Even though machine learning is gaining popularity in epidemiology, a recent systematic review showed no evidence of superior performance of machine learning over logistic regression for clinical prediction models, such as ours [34]. The most common methods of machine learning in the systematic review were random forests, artificial neural networks, and support vector machines [34]. The Area Under the Curve (AUC) of machine learning and logistic regression models for clinical risk prediction were reported to be similar when comparisons had a low risk of bias [34]. However, an important advantage of machine learning over logistic regression is in handling a huge number of predictors [35-37]. Moreover, machine learning is reported to work well for problems with a strong signal-to-noise ratio [38], that is gaming, writing recognition, electric forecasting, however clinical prediction problems tend to have poor signal to noise ratio [39]. Therefore, we are of the opinion that the use of the logistic regression sufficed.

Implications for clinical practice and research

Both in research and practice, a lot of emphasis has been placed on pain reduction in low back pain. However, the findings of *Chapter 2* suggest that the way an individual's pain influences his or her daily activities (disability) has more bearing on health-related quality of life and/or costs than pain reduction alone [13]. A systematic review by Clark et al. (2019), showed that surgery in adults with lumbar radiculopathy reduced pain meaningfully, but improvements in physical functioning and disability were small to trivial [40]. This finding is in line with that of *Chapters 2* and *7* and again suggests that we should focus more on disability instead of pain levels during treatment as that has more bearing on health-related quality of life, costs, and self-perceived recovery. However, the evidence from the randomized controlled trials involved in the systematic review by Clark et. al (2019) were graded as "low" to "very low" [40]. Hence, there is need for new high quality randomized controlled trials on this topic. In spite of this, the findings from *Chapter 2*, *Chapter 7*, and the systematic review suggest that a reduction of pain severity does not necessarily result in an improvement in disability and

that disability seems to have a bigger impact on health-related quality of life, costs, and self-perceived recovery than pain [13,40,41]. Therefore, it might be worthwhile for future research and interventions to focus more on what patients want and can do with their pain, instead of just trying to get rid of the pain. In clinical practice, clinicians can also focus their treatments on alleviating disability in this patient group and enabling patients to cope with their pain. Possible means to achieve this may include focusing on activities for daily living and returning to work if this is what a patient is interested in. In addition, the mechanisms causing the association between pain and disability are not well known [32]. Therefore, future research should prioritize understanding this relationship. Future research should also explore the barriers in the healthcare system that non-nationals/immigrants with low back pain face and how that contributes to poor outcomes and high societal costs.

THEME B: EFFECTIVENESS AND COST-EFFECTIVENESS OF SCIATICA TREATMENTS

Research questions:

1. Is combination therapy (MDT &TESIs) effective and cost-effective compared to usual care among sciatica patients with an indication for surgery? *(Chapter 4 and 5)*
2. Is exercise therapy for sciatica effective? *(Chapter 6)*

Summary of main findings:

Chapter 4 and *Chapter 5* describe the design and the preliminary (cost-)effectiveness analysis of the PLUS-study, which aimed to compare combination therapy, consisting of Mechanical Diagnosis and Treatment (MDT) and Transforaminal Epidural Steroid Injections (TESIS), to no intervention while being on the waiting list for lumbar herniated disc surgery (i.e. usual care). We intended to include 146 patients, but recruitment has been poor and we only managed to include 56 patients in this preliminary analysis. The patients were randomized to combination therapy (n= 27) and usual care (n=29). The primary outcome was the number of patients undergoing lumbar disc surgery during 6-month follow-up. The cost-effectiveness analysis was performed from the societal perspective. Results of this preliminary analysis suggest that combination therapy can prevent lumbar disc herniation surgery in comparison with usual care, that both strategies have similar clinical effects, and that the combination therapy's cost-effectiveness seems promising. The latter is based on the fact that the probability of combination therapy being cost-effective compared with usual care was 0.6 if decision-makers are not willing to pay anything for preventing a surgery, and that this probability increased with increasing values of willingness to pay to 1.0 at a threshold of €20,000/surgery prevented. However, we do not yet know what will happen if the follow-up duration is extended to 12 months and more patients are included. The latter is important, because the analyses were quite underpowered. That is, according to our sample size calculation, we needed to include 146 patients to detect a difference of 30% (i.e. a reduction of 30% of surgeries in the intervention group compared to usual care). However, we

have only managed to include 68 patients to date, of which only 56 were followed-up for 6 months and could be included in the preliminary analysis.

Chapter 6 describes a systematic review that summarizes the current evidence on the effectiveness of exercise therapy in patients with sciatica. Exercise is the most commonly used conservative treatment for sciatica in primary care. From the results, the effectiveness of exercise therapy on sciatica was unclear. Exercise had a small short-term effect on functional status compared to other therapies ([MD: -8.4; 95%CI:-15.70 to -1.10]; 0-100 scale), large short-term effect on pain compared to no therapy ([MD: -2.07; 95% CI: -3.24 to -0.89], 0-10 scale); and a medium long-term effect on GPE when given as an adjunct therapy ([RR: 1.42; 95% CI: 1.11 to 1.81]; RD: 0.23; NNT=4). Other comparisons had small, non-statistically significant differences. None of the studies compared exercises to surgery. Certainty of evidence was "low" to "very low" for all outcomes. This was due to the fact that the included studies had a large unexplained statistical heterogeneity and the effects could not be pooled. Sample sizes of the included studies were small and the quality of included studies was low.

Discussion

The preliminary results of the (cost-)effectiveness analysis *(Chapter 5)* indicate that combination therapy while being on the waiting list for lumbar herniated disc surgery can prevent such surgeries from happening. These results are quite promising, but we do not know what will happen to the results when 146 patients are included in the analysis compared to the 56 patients. The limitation of having a small sample size and slow recruitment, which threatened the study's statistical power and external validity, respectively, will be further discussed below. This is particularly important, because the results in *Chapter 6* regarding the effectiveness of exercise were also threatened by small sample sizes of the studies included in the systematic review. Thus, patient recruitment seems to be an issue many researchers are struggling with.

Low inclusion rate in randomized controlled trials

Slow recruitment and failure to reach the planned sample size within the planned timeframe and trial funding period is commonplace in randomized controlled trials [42-44]. Our randomized controlled trial *(Chapter 5)* is an example of such randomized controlled trials. Based on our sample size calculation [41] we set out to recruit 146 patients over a period of one year. However, 2.5 years later we have only managed to recruit 68 patients, 56 patients of which completed 6 months follow-up and were included in the preliminary (cost-)effectiveness analysis *(Chapter 5).* Another example of a discontinued randomized controlled trial, is of a trial that aimed to recruit (sub-)acute sciatica patients in general practice in the Netherlands, but was discontinued due to unsuccessful patient recruitment [45]. After 12 months, only 8 out of 234 patients were recruited [45]. General practitioners involved in patient recruitment of the randomized controlled trial cited low incidence rate, strict eligibility, strong patient/general practitioner preference, and time constraints as one of the main reasons for unsuccessful patient recruitment [45]. The PACE trial was also discontinued due to problems in patient recruitment in the general practice in the Netherlands [46]. After about 6 months, the trial recruited 4 out of the required 800 acute non-specific low back pain patients [46]. Possible explanations for the problems in recruitment for the PACE trial were categorized into patient factors (e.g. patients had different expectations), general practitioner factors (e.g. lack of time for the trial or trial was forgotten due to tasks) and research factors (e.g. research logistics disturbed usual clinical care) [46]. A problem with insufficient patient recruitment is that research questions addressed in the trial remain relevant, but unanswered. Another problem is the wastage of research resources, especially when there is no publication of the results and/or the reasons for the trial's premature discontinuation [47].

The PLUS-study was not a general practitioner trial, but a trial in surgical clinics. Randomized controlled trials performed in surgery clinics typically experience even more difficulty in patient recruitment than general practitioner trials [48], threatening the validity of their findings [49]. In the systematic review of Abraham et al. (2006), surgeons were asked why they did not want to recruit eligible patients into surgical trials.

Their reasons included; patients having strong preferences for a particular therapy, patients not wanting to be randomized, patients' fear of receiving negative outcomes from a therapy patients deemed inferior, and difficulty in understanding/following the trial procedure. Other possible explanations for difficulty in recruitment in surgical trials include the fact that only a small sub-group of the total population has an indication of surgery, presence of comorbidities that prevent surgery and resist inclusion, and patient's/surgeon's strong preferences for a certain treatment option [48]. Difficulty in recruitment is also particularly observed in diseases or conditions with high incidence and prevalence rates [48]. This is because not all patients will be eligible, but only a sub-group [48]. The PLUS-study dealt with low back pain patients with an indication for lumbar herniated disc surgery. A condition with high prevalence rates in the Netherlands and worldwide. Issues that possibly affected recruitment for the PLUS-study are classified into surgeon factors, research factors, and patient factors, and will be further discussed below.

Surgeon factors

1. *Surgeon as the primary recruiter*

In retrospect, one of the most important challenges of the PLUS-study was having the neurosurgeons as the sole recruiters of patients. Recruiting patients is particularly challenging for neurosurgeons, because of their high workload. We initially opted for this strategy because only neurosurgeons can indicate whether a patient is a surgical candidate or not. This is important in our study, because the primary goal of the combination therapy was to reduce the number of surgeries in sciatica patients with an indication for surgery. Research assistants were employed in an attempt to lessen the burden on neurosurgeons and improve recruitment, but inclusion rates remained low. In addition, neurologists could have been part of the patient recruitment process, as they saw hernia patients before referring them to the neurosurgeons or for epidural injections, hence providing an opportunity to identify potential patients before they received epidural injections or were referred to one of the participating centers. Recruiting patients before they received epidural

injections was important, because having received an epidural injection in the previous 6 months was one of the exclusion criterion [41]. During the course of the study, some attempts were made to include neurologists in the recruitment process. These attempts were without success, because a neurosurgeon was still required to confirm whether a patient was a surgical candidate or not and some changes were required in the day-to-day functioning structures of the participating centers regarding referrals between departments and agendas. On the other hand, having the neurosurgeon as the primary recruiter was advantageous, because the relationship between clinicians and patients is significant for engaging patient as participants in research [49]. This is also evidenced by the relatively high number of patients with complete data in the PLUS-study. Moreover, surgical consultants are reported to be more successful in patient recruitment compared to other members of staff, because patients either feel unable to refuse their invitation or have confidence in the research when it is supported by them [50]. The manner in which research is communicated to patients can also influence patients' acceptance of equipoise (i.e. that it is uncertain which of the interventions is superior over the other) [49,51,52]. Patients tend to trust and value the opinions of their doctors. In addition, senior staff, such as neurosurgeons, have networks established, which they can utilize to recruit patients. The opinions/preferences of the neurosurgeons could also influence patient recruitment negatively, for example, if neurosurgeons have a strong preference for one intervention over the other [50]. Surgeons usually struggle with equipoise when there is patient randomization to a less favored technique or intervention [53]. The lack of equipoise is one of the important reasons why there are relatively few successful randomized controlled trials involving surgical interventions [54-56]. Usually, surgeons use a particular procedure, because they believe it gives better outcomes [53]. Hence, they hope that randomized controlled trials testing the outcomes of that procedure sustain those beliefs [53]. In addition, surgical consultation is intended to justify choices taking into account the benefits, risks, and patient's wishes [57]. Surgeons might feel uncomfortable expressing uncertainty in a conversation were the role of the surgeon is to assist in decision-making,

install confidence, and trust [57]. Furthermore, doctors participating in randomized controlled trials are reported to suffer emotionally and intellectually due to conflicts between their clinical instincts, the needs of randomized controlled trials, treatment preferences and worries about patient eligibility and safety [54]. This discomfort affects their readiness to recruit patients for randomized controlled trials. A proposed solution includes separation of surgical consultation and randomization meetings [57]. This was the case in the PLUS-study, in which randomization was conducted by an independent researcher [41]. Another possible solution is to conduct randomization prior to surgical consultation, allowing the uncertainty to be resolved, then the surgeon can discuss the implications of the assigned intervention [57]. However, this was not possible in the PLUS-study, because the neurosurgeon played a key role in the recruitment of eligible patients and randomization took place after potential patients were identified as candidates for surgery and recruited. Training and support for doctors included in the recruitment process can help them to engage easily with patients, become familiar and comfortable with randomized controlled trial concepts such as equipoise and participate in a more resilient recruitment process [54].

2. Organizational context on recruitment

Organizational norms, structures, and processes can serve as significant barriers or facilitators in the recruitment of patients as research subjects [50]. Waiting lists of longer than four weeks in the participating hospitals/centers facilitated patient recruitment for the PLUS-study. However, too long waiting lists became barriers to patient recruitment, because potential patients were then referred to other hospitals/centers or received pain injections. Too short waiting lists were also a barrier to patient recruitment, as they did not leave enough time for a patient to receive the intervention (combination therapy) if randomized to the intervention arm of the PLUS-study. In addition, after waiting for a long time to see a neurosurgeon, patients were likely to choose for surgery and it also became difficult for neurosurgeons to deny their requests. If conducting research is imbedded in the structures and culture of organizations, patient recruitment might improve. This is because organizations can

reduce workload for clinicians as an incentive to improve patient recruitment [50] and improve cohesion between different disciplines. In case of the PLUS-study this might have been achieved by involving not only the neurosurgeons but also the hospital management in the designing of the study.

Research factors

1. Number of recruitment centers and location

Initially, we recruited five participating hospitals. This was based on the amount of lumbar herniated disc patients expected to visit the participating hospitals. However, when the rate of inclusion remained low, because some of the patients were not asked and others did not meet the inclusion criteria or did not want to participate, additional hospitals were asked to participate in the study. The distance of the hospital from one of the four outpatient clinics (Rugpoli's) where the combination therapy was offered influenced which hospitals were asked to participate. Although the addition of two participating hospitals in the PLUS-study improved patient recruitment to some extent, the targeted sample size was not reached. Recruiting a larger number of participating hospitals could have helped optimize patient recruitment. However, inclusion of more hospitals was further limited by the fact that many of them were located too far away from the outpatient clinics. Setting up MDT plus TESIS intervention within the participating hospitals was therefore attempted, but this was not successful.

2. Funding and time in randomized controlled trials

The structure of funding, including the duration of funding, may influence the success of a clinical trial. The time frame for completing the PLUS-study was three years, whereas the median duration of a randomized controlled trial from enrollment to publication is five and half years [58]. Thus, the duration of the PLUS-study might have been too short to begin with. Whether additional time to the PLUS-study would have improved patient recruitment, is uncertain. This is because more factors, as mentioned above, contribute to the success of a randomized controlled trial. Nonetheless, more time would have enabled recruitment of additional participating hospitals, which might

have in turn improved patient inclusion. To prevent potential wastage of public resources due to incomplete trials, grant organization could ensure that funding time of clinical trials are in line with the average time it takes to complete a randomized controlled trial. In addition, the trial extension process might be made easier. It would also be good to make leeway in the time allocated to setting up a trial. This is because the process of acquiring approval from the relevant authorities might not always be smooth sailing and the time frame varies depending on the authority in question and whether or not the approval is granted upon initial request.

Patient factors

1. Patients preference/randomization/consent

Patient preference for a particular intervention is seen as one of the most important reasons for poor inclusion in randomized controlled trials [49]. Randomization, a key design aspect of a randomized controlled trial, requires a state of indifference, i.e. personal equipose [59], and it evades choice [57]. As a result, patients who believe in the superiority of one intervention over the other typically experience difficulty with being randomized. Studies report that a patient's belief in clinical equipoise, i.e. that there is not one intervention better than the other, is key to a patient's consent to randomization [51]. Thus, patients may be more willing to participate in randomized controlled trials, such as the PLUS-study, if they understand and accept equipoise [51]. However, patients, like clinicians have ideas, experiences, and preferences about the interventions under investigation, which can negatively affect recruitment [59]. Some patients in the PLUS-study also refused participation because they wanted to remain on the waiting list for surgery and did not want to be randomized.

Recruiters' own views may contribute to patient preferences [60]. When recruiters carefully explore these preferences and provide information, only a minority of patients remain committed to their initial preference [60]. However, some recruiters are reluctant to explore patient preferences, particularly when they concur with their own views [60]. In such circumstances, they reason that exploring the patient's

preferences will be too coercive [60]. However, patient beliefs and preferences are explored during recruitment meetings to ensure that the patient makes an informed decision [59]. A failure to do so can impair the consent process [59]. Whether and when such an exploration is too coercive is unclear and requires further research to understand were to draw the line. Recruitment meetings for the PLUS-study took place between the patient and the neurosurgeon. Hence, the neurosurgeons decided on which patients to ask to participate in the study, and informed the patient about the study. As a consequence, whether patients that chose to participate in the study had clinical equipoise is unknown and neither is the impact of the view of the neurosurgeons on the participation of patients. Further research should explore the balance between coercion and patient preferences without dispelling any misinformation.

2. Burden

Patients have also been found to refuse to participate in clinical trials if they think that the trial is too complex [49] or that participation will require a lot of their time. The PLUS-study had seven follow-up moments. Even though we aimed to make questionnaires as simple and less time consuming as possible, patients may have still found it too involving to complete questionnaires. Patients also received reminders and were offered support in completing their questionnaires if need be. In spite of these measures being in place, many patients still refused participation, some of which cited time constraints as one of the main reasons.

Generalizability

Recruitment rate of patients in the PLUS-study was slow, therefore the generalizability of the results to other Dutch low back pain patients is limited. Hence, it is uncertain whether the included participants in *Chapter 5*, are a true representation of all patients waiting for lumbar herniated disc surgery in the Netherlands. The difference can be expected because a relatively small section of eligible patients eventually participated. To illustrate, only 7% of the patients on the waiting list for surgery in the PLUS-study had a low level of education. Whereas, van Dongen et. al (2017) reported a significantly higher total of

72% of patients referred to surgery as having a low level of education [18]. The generalizability of the cost-effectiveness analysis *(Chapter 5)* to other countries may be limited due to differences in healthcare and social wealth systems, since this analysis pertains to a Dutch context. In addition, our economic evaluation took both a healthcare and a societal perspective which is not necessary the case in other economic evaluations. In the United Kingdom, for example, the National Institute for Clinical Excellence (NICE) recommends that costs should be measured from the perspective of the National Health Service and Personal Social Services (NHS & PSS) [61]. The selection process of the centers and clinicians may have also had important implications for the study's external validity [62]. In the PLUS-study, the participating hospitals were selected based on their close proximity to the outpatient clinics were the intervention was given, whereas other hospitals that might have the potential to recruit more patients were excluded.

Implications for clinical practice and research

The preliminary analysis for the PLUS-study in *Chapter 5* indicates that combination therapy while being on the waiting list for lumbar herniated disc surgery can prevent such surgeries from happening compared with usual care, that both have similar clinical effects, and that the combination therapy's cost-effectiveness seems promising. This is an important and promising finding, because refraining patients from expensive, invasive interventions in secondary care and treating them in primary care seems efficient and safe. Since no significant differences regarding physical functioning, pain and perceived recovery (secondary outcomes) between the intervention and control group were observed, the main effect of the combination therapy on prevention of surgery, might be triggered by other intervention-related factors, such as the provision of information regarding the particular condition, self-assurance or validation of patient complaints by the therapists/physician. Hence, in clinical practice, it is important for the clinicians to pay extra attention to such patients, validate their complaints and offer advice regarding self-management of their complaints.

Research indicates that multiple factors contribute to unsuccessful recruitment of patients in randomized controlled

trials. Therefore, it is important to address these factors from all angles included. That is, from the clinicians (i.e. surgeons), patients, organizations, and researchers' point of view. Researchers should aim to minimize the burden on clinicians participating in studies by minimizing interference into day-to-day practice and providing protocols that are easy and simple to follow. Recruiting patients for research should not be cumbersome for the clinicians/specialists and should not involve any additional paperwork, but utilize existing referral systems [45]. Researchers should also publish articles of failed randomized controlled trials and the reasons contributing to their failure. Publication of such articles, is informative, as other researchers become aware that the relevant questions that the trial intended to answer have not been answered as yet. In addition, this will give pointers pertaining to the causes of failure, which can be valuable to researchers conducting similar studies.

Training and support should be provided for doctors/specialists included in the recruitment process. This is important because doctors participating in randomized controlled trials are reported to suffer emotional and intellectual conflicts between their clinical instincts, needs of randomized controlled trials, treatment preferences, and worries about patient eligibility and safety, consequently affecting recruitment negatively [54]. Training can help them to engage easily with patients, become familiar and comfortable with randomized controlled trial concepts, such as equipoise, and participate in a more resilient recruitment process [54].

Conducting research should be imbedded in the structures and culture of healthcare organizations, to promote evidence-based practice and to value research. Organizations are in a position to reduce workload for clinicians as an incentive to improve patient recruitment [50], improve working together between different disciplines and also train employees in research activities. Organizations/hospitals should also foster relations with academic institutions and develop a research agenda with interests that are relevant to both parties.

A patient-centered approach in medicine is gaining momentum as patients assume more active roles in their own medical decisions [63]. However, the same cannot be said for clinical research, which is still carried out on patients [63], whereas the

patient-centered approach is not viable without patients having a key role in care decisions and in the research that influences these decisions [64]. The Netherlands Organisation for Health Research and Development (ZonMW), is paying more and more attention to the participation of patients in research project groups. This in itself is a positive development. The various perspectives that patients hold regarding their health conditions should be used to inform research, which may in turn inform decision-making [64]. Various ways to engage patients in clinical research and their impact on inclusion rates in randomized controlled trials should be explored. ZonMw reports that the societal value of research increases when patients are involved in research [65]. In addition, reaching the intended patient group in a study can be improved when patients and researchers collaborate [65]. Clinical trials should therefore become more accessible to patients to improve the rate of inclusion. Usually, patients whose doctor/clinician is the principal investigator in a clinical trial are eligible to participate in a study. Potential patients seen by other doctors/clinicians might not be eligible to participate in a study if their doctor/clinician is not a principal investigator. Hence potential participants seen by other clinicians may be omitted in the trial. Further research should also explore the balance between coercion and patient preferences without dispelling any misinformation, to further understand to what extend patient preferences can be influenced by recruiters. In addition, researchers should explore the impact of a research conscious society on the participation of patients in clinical trials.

THEME C: METHODOLOGICAL STUDIES

Research questions:

1. To what extend is the association between the Global Perceived Effect (GPE) and change in pain and functional status influenced by current health status? *(Chapter 7)*
2. Does the statistical approach in trial-based economic evaluations matter? *(Chapter 8)*

Summary of main findings:

Chapter 7 investigated the construct validity of the GPE scale for measuring self-perceived recovery in patients with sciatica. The results showed that GPE was statistically significantly associated with change in leg pain (OR:1.04;95%CI:1.02-1.05), change in back pain (OR:1.02;95%CI:1.01-1.04), and change in functional status (OR:1.08;95%CI:1.04-1.12). Adding current pain and functional status, respectively, decreased the magnitude of the associations and increased the models' explained variance. The effect of adjusting for current pain or functional status did not systematically decrease with longer follow-ups durations. Self-perceived recovery had the strongest association with change in functional status (OR:1.25;95%CI:1.10-1.42) when GPE was regressed upon pain and functional status simultaneously. To conclude on the results, GPE does not seem to be a true transition scale for patients with sciatica, as current health status considerably impacts GPE scores. Moreover, in contrast to other studies, this does not seem to depend on follow-up duration.

Chapter 8 evaluated whether accounting for missing data, skewed costs, the correlation between costs and effects, and baseline imbalances in trial-based economic evaluations has an impact on the results. The results showed that failure to adequately account for these statistical challenges in trial-based economic evaluations may affect cost-effectiveness results substantially. Therefore, when conducting trial-based economic evaluations, it is of utmost importance to first check the data to identify statistical challenges that need to be accounted for in the analysis, and then adequately deal with them. Furthermore, it is worthwhile to develop consensus among researchers about frameworks and guidelines on how to best analyze trial-based economic evaluations.

DISCUSSION

Global Perceived Effect as a true transition scale

Various studies including our own *(Chapter 7)* have shown that GPE is highly influenced by a patient's current health status and is therefore not a "true transition scale" when assessing recovery. This has been demonstrated in different populations, including sciatica patients. In addition, our study *(Chapter 7)* indicates that the GPE does not seem to perform adequately as a transition scale, regardless of the transition period. In contrast, the studies of Kamper et al., and Schmitt and Fabio (2005) and Kamper et al (2010) concluded that GPE scales performed adequately over short transition periods and were less accurate with longer recall periods. However, Kamper et al. (2010) only included sciatica patients who suffered substantial residual complaints six weeks after lumbar disc surgery. This is contrast to our population, which included all post-surgery sciatica patients regardless of whether they had a referral for early rehabilitation post surgery. It is highly likely that sciatica patients who suffer substantial residual complaints are more alert on the changes they experience towards recovery compared to other post-operative patients that do not experience substantial residual complaints. Hence, the patient population of Kamper et. al (2010), is not representative of the general sciatica post-operative population. This could also explain why in our study *(chapter 7)* GPE scales did not perform adequately as a transition scale, regardless of transition period, and these results were different from that of Kamper et. al (2010).

In *Chapter 8*, one could argue that the t-test approaches (one that did not correct for any of the statistical challenges) which are presented are rarely used by analysts in practice and so is the use of mean imputation to handle missing outcome data. However, this might be counterintuitive, because research does indicate that many researchers conducting trial based economic evaluations do not do this. In their systematic review El Alili et al. (2017) reported that the majority of cost-effectiveness evaluations do not comply with the recommended statistical methods and reporting guidelines [66]. Moreover, national pharmacoeconomic guidelines were reported to provide no or little guidelines on how statistical challenges in trial-based

economic evaluations can be addressed [67].

Another critic on *Chapter 8* could be the chosen method of dealing with the statistical challenges. That is, some other, important alternative methods such as different parametric distributions to handle skewness, hurdle or two-part regressions to handle spikes in the empirical distributions, alternative types of multiple imputation methods, were not mentioned and/or assessed. Secondly, the statistical challenges chosen might not cover the whole selection of statistical challenges encountered in trial-based economic evaluations. However, it is necessary to note that the statistical challenges identified in this study are only a selection of possible statistical issues that might arise when analyzing trial-based economic evaluations. In addition, statistical methods are continuously developing. Hence, in order to improve the statistical quality of trial-based economic evaluations, expanding on the health economic literature showing the current practice might be beneficial. No guidance regarding the statistical methods to use in trial based economic evaluations are provided in the current national pharmacoeconomic guidelines [67].

Generalizability of our results

In *Chapter 7*, the generalizability of the results to other populations beyond postoperative sciatica patients, included in our study, might be limited. The data used in *Chapter 7* was from a randomised controlled trial and the participants might be slightly different to sciatica patients in general. Nonetheless, the findings of *Chapter 7* add to the existing evidence regarding the construct validity of the GPE as specific focus was placed on a specific patient group in the musculoskeletal domain.

In *Chapter 8*, the generalizabilty is limited because the findings depend on the characteristics of datasets analyzed. Only two randomised controlled trials, one in back pain and the other in diabetes were used. Therefore, it is unknown whether the finding that that failure to adequately account for baseline imbalances, skewed costs, correlated costs and effects, and missing data in trial-based economic evaluations may have a substantial impact on cost-effectiveness results, is applicable to all other trial-based economic evaluations. However, the main message from *Chapter 8* is that it is of paramount importance to

align the statistical approach with the statistical challenges of a particular dataset, which is highly likely to be relevant to all trial-based economic evaluations. Further research could explore the performance of the various approaches using simulated data [68].

Implications for clinical practice and research:
GPE is highly influenced by a patient's current health status, even at short-term follow-up, and can therefore not be considered a true transition scale. Hence, further research into an alternative scale for the GPE is warranted as well as research into exploring whether the GPE is indeed influenced more by functional status than by pain as our results in *Chapter 7* suggest. Since the GPE is highly influenced by current health status, clinicians and researchers should bare in mind that the patient's improvement, or lack of thereof, is influenced by how the patient feels at the time of asking. Moreover, clinicians can calculate the change scores themselves by noting the baseline scores and follow-up scores for (e.g. pain, function) for themselves. Changing the instrument used to measure recovery or developing a new instrument might solve this problem.

Chapter 8 touches on some of the possible statistical challenges and possible statistical approaches/methods to handle them. However, the statistical methods are constantly being improved, and hence it becomes of value to expand the literature in health economics showing the current state of affairs. In addition, development of guidelines and frameworks in which specific statistical methods are recommended to be used in trial-based economic evaluations might further improve the statistical quality of trial-based economic evaluations. Furthermore, it is worthwhile to develop consensus among researchers about frameworks and guidelines on how to best analyze trial-based economic evaluations.

Concluding remarks:

The present thesis indicated that:

- Even though there is a lot of emphasis on pain in low back pain research, pain-related disability in low back pain seems

to have a lot more impact on societal costs and health-related quality of life. Therefore, future research and interventions in low back pain should focus more on disability.

- Disability in low back pain results in higher societal costs compared to pain. Yet, the mechanisms causing the association between pain and disability are not well understood. Neither is the possible mediating role of disability in the relationship between pain and costs and health-related quality of life. Both issues warrant further exploration.
- Poor physical health, high functional disability, low health-related quality of life, high impact of pain experience, non Dutch nationality and decreasing pain were found to be the most robust predictors of high societal costs. Thus, these factors have important implications in understanding high cost users in low back pain and tailor effective cost reduction interventions.
- Based on our systematic review there was no consistent effect of exercise for the treatment of sciatica. Thus, better quality studies in this area are required.
- A combination therapy of MDT and TESIS while being on the waiting list for lumbar disc herniation surgery can prevent such surgeries from happening in comparison with being on the waiting list alone (i.e. usual care), both strategies have similar clinical effects, and the combination therapy's cost-effectiveness seems promising. However, we do not yet know what will happen if the follow-up duration is extended to 12 months and more patients are included.
- The main effect of the combination therapy of MDT and TESIS on preventing surgery was not associated with improved clinical outcomes compared with usual care. The combination therapy's effect on surgery rate might be triggered by other intervention-related factors, such as the provision of information regarding the particular condition, self-assurance, and validation of patient complaints by the therapists/physician.
- Multiple factors contribute to unsuccessful recruitment of patients in randomized controlled trials. Therefore, it is important to address these factors from all angles included. That is, from the clinicians (i.e. surgeons), patients,

organizations and researchers' point of view.

- GPE is highly influenced by a patient's current health status, even at short-term follow-up, and can therefore not be considered a true transition scale. Hence, further research into an alternative scale for the GPE is warranted.
- When conducting a trial-based economic evaluation, it is of paramount importance to know your dataset and to align the statistical approaches with the statistical challenges of a dataset, that is adequately accounting for baseline imbalances, skewed costs, correlated costs and effects, and missing data. Ignoring stastical challenges/not adequately accounting for them may lead to biased cost-effectiveness results.

REFERENCES

1. Lambeek LC, van Tulder MW, Swinkels IC, Koppes LL, Anema JR, van Mechelen W. The trend in total cost of back pain in The Netherlands in the period 2002 to 2007. *Spine*. 2011;36(13):1050-1058.

2. Dieleman JL, Baral R, Birger M, Bui AL, Bulchis A, Chapin A, Hamavid H, Horst C, Johnson EK, Joseph J, Lavado R, Lomsadze L, Reynolds A, Squires E, Campbell M, DeCenso B, Dicker D, Flaxman AD, Gabert R, Highfill T, Naghavi M, Nightingale N, Templin T, Tobias MI, Vos T, Murray CJ. US Spending on Personal Health Care and Public Health, 1996-2013. *JAMA*. 2016; 316(24):2627-2646.

3. Crow WT, Willis DR. Estimating cost of care for patients with acute low back pain: a retrospective review of patient records. *J Am Osteopath Assoc*. 2009 (4):229-33.

4. Walker B, Muller R, Grant W. Low back pain in Australian adults: the economic burden. *Asia Pacific Journal of Public Health*. 2003;15(2):79-87.

5. Wieser S, Horisberger B, Schmidhauser S, Eisenring C, Brügger U, Ruckstuhl A, Dietrich J, Mannion AF, Elfering A, Tamcan O, Müller U. Cost of low back pain in Switzerland in 2005. *Eur J Health Econ*. 2011; 12(5):455-67.

6. Maniadakis N GA. The economic burden of back pain in the UK. *Pain*. 2000;84(1):95-103.

7. Maniadakis N, Gray A. The economic burden of back pain in the UK. *Pain*. 2000;84(1):95-103.

8. Tsuboi Y, Murata S, Naruse F, Ono R. Association between pain-related fear and presenteeism among eldercare workers with low back pain. *Eur J Pain*. 2019 Mar;23(3):495-502.

9. Sadosky AB, DiBonaventura M, Cappelleri JC, Ebata N, Fujii

K. The association between lower back pain and health status, work productivity, and health care resource use in Japan. *Journal of pain research*. 2015;8:119.

10. Ostelo RW, de Vet HC. Clinically important outcomes in low back pain. *Best Pract Res Clin Rheumatol*. 2005;19(4):593-607.

11. Luo N, Johnson J, Coons SJ. Using instrument-defined health state transitions to estimate minimally important differences for four preference-based health-related quality of life instruments. *Med Care*. 2010 Apr;48(4):365-71.

12. Walters SJ, Brazier JE. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. *Qual Life Res*. 2005;14(6):1523-32.

13. Mutubuki EN, Beljon Y, Maas ET, Huygen FJPM, Ostelo RWJG, van Tulder MW, van Dongen JM. The longitudinal relationships between pain severity and disability versus health-related quality of life and costs among chronic low back pain patients. *Qual Life Res*. 2020 Jan;29(1):275-287.

14. Bewick V, Cheek L, Ball J. Statistics review 14: Logistic regression. *Crit Care*. 2005;9(1):112-118.

15. Greiner M, Pfeiffer D, Smith RD. Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. *Prev Vet Med*. 2000 May 30;45(1-2):23-41.

16. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, Pencina MJ, Kattan MW. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology*. 2010;21(1):128-38.

17. Maas ET, Juch JN, Groeneweg JG, et al. Cost-effectiveness of minimal interventional procedures for chronic mechanical low back pain: design of four randomised controlled trials with an economic evaluation. *BMC Musculoskelet Disord*. 2012;13:260.

18. van Dongen JM, van Hooff ML, Spruit M, de Kleuver M, Ostelo

RWJG. Which patient-reported factors predict referral to spinal surgery? A cohort study among 4987 chronic low back pain patients. *Eur Spine J*. 2017;2782–2788.

19. Mutubuki EN, Luitjens MA, Maas ET, Huygen FJPM, Ostelo RWJG, van Tulder MW, van Dongen JM. Predictive factors of high societal costs among chronic low back pain patients. *Eur J Pain*. 2020;24(2):325-337.

20. Messerer D, Porzsolt F, Hasford J, Neiss A. Advantages and problems of multicenter therapy studies exemplified by a study of the treatment of metastasizing renal cell carcinoma with recombinant interferon-alpha-2c. *Onkologie*. 1987 Feb;10(1):43-9.

21. Youssef N, Reinhart K, Sakr Y. The pros and cons of multicentre studies. *Neth J Crit Care*. 2008;12(3).

22. Brzoska P, Voigtländer S, Spallek J, Razum O. Utilization and effectiveness of medical rehabilitation in foreign nationals residing in Germany. *European journal of epidemiology*. 2010;25(9):651-660.

23.Brzoska P. Disparities in health care outcomes between immigrants and the majority population in Germany: A trend analysis, 2006–2014. *Plos one*. 2018;13(1).

24. Cano A, Mayo A, Ventimiglia, M. Coping, pain severity, interference, and disability: the potential mediating and moderating roles of race and education. *The Journal of Pain*. 2006;7(7):459-468.

25. Edwards RR, Moric M, Husfeldt B, Buvanendran A, Ivankovich O. Ethnic similarities and differences in the chronic pain experience: a comparison of African American, Hispanic, and white patients. *Pain Medicine*. 2005;6(1):88-98.

26. Carey TS, Garrett JM. The relation of race to outcomes and the use of health care services for acute low back pain. *Spine*. 2003;28(4):390-394.

27. Kreling MCGD, de Mattos Pimenta CA, Garanhani ML. Racial discrimination in pain management. *Revista Dor*. 2014;15(3).

28. Fink-Miller EL, Long DM, Gross RT. Comparing chronic pain treatment seekers in primary care versus tertiary care settings. *The Journal of the American Board of Family Medicine*. 2014;27(5): 594-601.

29. Bennet JA. Focus on research methods mediator and moderator variables. Nursing Research: Conceptual and Statistical Differences. *Research in Nursing & Health*. 2000;23:415-420.

30. Zale EL, Ditre JW. Pain-Related Fear, Disability, and the Fear-Avoidance Model of Chronic Pain. *Current opinion in psychology*. 2015;5:24–30.

31. Lee H, Hübscher M, Moseley GL, Kamper SJ, Traeger AC, Mansell G, McAuley JH. How does pain lead to disability? A systematic review and meta-analysis of mediation studies in people with back and neck pain. *Pain*. 2015;156(6):988-97.

32. Lee H, Hübscher M, Moseley GL, Kamper SJ, Traeger AC, Mansell G, McAuley JH. How does pain lead to disability? A systematic review and meta-analysis of mediation studies in people with back and neck pain. *Pain*. 2015;156(6):988-97.

33.Twisk J. *Applied Longitudinal Data Analysis for Epidemiology: A Practical Guide*. Cambridge: Cambridge University Press; 2013.

34. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol*. 2019;110:12-22.

35.Kononenko I. Machine learning for medical diagnosis: history, state of the art and perspective. *Artificial Intelligence in medicine*. 2001;23(1):89-109.

36. Goldstein BA, Navar AM, Carter RE. Moving beyond

regression techniques in cardiovascular risk prediction: applying machine learning to address analytic challenges. *European heart journal*. 2017;38(23):1805-1814.

37. Luo W, Phung , Tran T, Gupta S, Rana S, Karmakar C, Venkatesh S. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. *Journal of medical Internet research*. 2016;18(12).

38.Mitchell TM. Does machine learning really work? *AI magazine*. 1997;18(3):11-20.

39. Ennis M, Hinton G, Naylor D, Revow M, Tibshirani R. A comparison of statistical learning methods on the GUSTO database. *Statistics in medicine*. 1998;17(21):2501-2508.

40. Clark R, Weber RP, Kahwati L. Surgical Management of Lumbar Radiculopathy: a Systematic Review. *Journal of general internal medicine*. 2019:1-10.

41. Mutubuki EN, van Helvoirt H, van Dongen JM, Vleggeert-Lankamp CLA, Huygen FJPM, van Tulder MW, Klopper-Kes HAHJ, Ostelo RWJG. Cost-effectiveness of combination therapy (Mechanical Diagnosis and Treatment and Transforaminal Epidural Steroid Injections) among patients with an indication for a Lumbar Herniated Disc surgery: Protocol of a randomized controlled trial. *Physiotherapy Research International*. 2020 ;25(1):e1796.

42. Al-Shahi Salman R, Beller E, Kagan J, Hemminki E, Phillips RS, Savulescu J, Macleod M, Wisely J, Chalmers I. Increasing value and reducing waste in biomedical research regulation and management. *Lancet*. 2014;383(9912):176-85.

43. McDonald AM, Knight RC, Campbell MK, Entwistle VA, Grant AM, Cook JA, Snowdon C. What influences recruitment to randomised controlled trials? A review of trials funded by two UK funding agencies. *Trials*. 2006;7(1):9.

44. Walters SJ, dos Anjos Henriques-Cadby IB, Bortolami O, Flight

L, Hind D, Jacques RM, Julious S A. Recruitment and retention of participants in randomised controlled trials: a review of trials funded and published by the United Kingdom Health Technology Assessment Programme. *BMJ open*. 2017;7(3):e015276.

45. van der Gaag WH, van den Berg R, Koes BW, Bohnen AM, Hazen LM, Peul WC, Voogt L, Verhagen AP, Bierma-Zeinstra SM, Luijsterburg PA. Discontinuation of a randomised controlled trial in general practice due to unsuccessful patient recruitment. *BJGP Open*. 2017;1(3)

46. Schreijenberg M, Luijsterburg PAJ, Van Trier YDM, Rizopoulos D, Koopmanschap MA, Voogt L, Maher CG, Koes BW. Discontinuation of the PACE Plus trial: problems in patient recruitment in general practice. *BMC Musculoskeletal Disorders*. 2018;19(1):146

47. Briel M, Olu KK, von Elm E, Kasenda B, Alturki R, Agarwal A, Bhatnagar N, Schandelmaier S. A systematic review of discontinued trials suggested that most reasons for recruitment failure were preventable. *J Clin Epidemiol*. 2016;80:8-15.

48.McLeod RS. Issues in surgical randomized controlled trials. *World journal of surgery*. 1999;23(12):1210-1214.

49. Abraham NS, Young JM, Solomon MJ. A systematic review of reasons for nonentry of eligible patients into surgical randomized controlled trials. *Surgery*. 2006;139(4):469-83.

50. Adams M, Caffrey L, McKevitt C. Barriers and opportunities for enhancing patient recruitment and retention in clinical research: findings from an interview study in an NHS academic health science centre. *Health Res Policy Syst*. 2015 12;13:8.

51. Mills N, Donovan JL, Smith M, Jacoby A, Neal DE, Hamdy FC. Perceptions of equipoise are crucial to trial participation: a qualitative study of men in the ProtecT study. *Control Clin Trials*. 2003 Jun; (3):272-82.

52. Paramasivan S, Huddart R, Hall E, Lewis R, Birtle A, Donovan

JL. Key issues in recruitment to randomised controlled trials with very different interventions: a qualitative investigation of recruitment to the SPARE trial (CRUK/07/011). *Trials*. 2011 Mar 15;12:78.

53. Devereaux PJ, Bhandari M, Clarke M, Montori VM, Cook DJ, Yusuf S, Sackett DL, Cinà CS, Walter SD, Haynes B, Schünemann HJ, Norman GR, Guyatt GH. Need for expertise based randomised controlled trials. *BMJ*. 2005;330(7482):88.

54. Donovan JL, de Salis I, Toerien M, Paramasivan S, Hamdy FC, Blazeby JM. The intellectual challenges and emotional consequences of equipoise contributed to the fragility of recruitment in six randomized controlled trials. *J Clin Epidemiol*. 2014 Aug;67(8):912-20.

55. McCulloch P, Altman DG, Campbell WB, Flum DR, Glasziou P, Marshall JC, Nicholl J; Balliol Collaboration, Aronson JK, Barkun JS, Blazeby JM, Boutron IC, Campbell WB, Clavien PA, Cook JA, Ergina PL, Feldman LS, Flum DR, Maddern GJ, Nicholl J, Reeves BC, Seiler CM, Strasberg SM, Meakins JL, Ashby D, Black N, Bunker J, Burton M, Campbell M, Chalkidou K, Chalmers I, de Leval M, Deeks J, Ergina PL, Grant A, Gray M, Greenhalgh R, Jenicek M, Kehoe S, Lilford R, Littlejohns P, Loke Y, Madhock R, McPherson K, Meakins J, Rothwell P, Summerskill B, Taggart D, Tekkis P, Thompson M, Treasure T, Trohler U, Vandenbroucke J. No surgical innovation without evaluation: the IDEAL recommendations. *Lancet*. 2009;374(9695):1105-12.

56. Ergina PL, Cook JA, Blazeby JM, Boutron I, Clavien PA, Reeves BC, Seiler CM; Balliol Collaboration, Altman DG, Aronson JK, Barkun JS, Campbell WB, Cook JA, Feldman LS, Flum DR, Glasziou P, Maddern GJ, Marshall JC, McCulloch P, Nicholl J, Strasberg SM, Meakins JL, Ashby D, Black N, Bunker J, Burton M, Campbell M, Chalkidou K, Chalmers I, de Leval M, Deeks J, Grant A, Gray M, Greenhalgh R, Jenicek M, Kehoe S, Lilford R, Littlejohns P, Loke Y, Madhock R, McPherson K, Rothwell P, Summerskill B, Taggart D, Tekkis P, Thompson M, Treasure T, Trohler U, Vandenbroucke J. Challenges in evaluating surgical innovation. *Lancet*. 2009;374(9695):1097-104.

57. Treasure T, Baum M. An approach to randomization into surgical clinical trials. *British Journal of Surgery*. 2017;104(1):11-12.

58. Ioannidis JP. Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. *Jama*. 1998;279(4):281-286.

59. Tomlin Z, deSalis I, Toerien M, Donovan JL. Patient advocacy and patient centredness in participant recruitment to randomized-controlled trials: implications for informed consent. *Health Expect*. 2014;17(5):670-82.

60. Donovan JL, Paramasivan S, de Salis I, Toerien M. Clear obstacles and hidden challenges: understanding recruiter perspectives in six pragmatic randomised controlled trials. *Trials*. 2014 Jan 6;15:5.

61. Massetti M, Aballéa S, Videau Y, Rémuzat C, Roïz J, Toumi M. A comparison of HAS & NICE guidelines for the economic evaluation of health technologies in the context of their respective national health care systems and cultural environments. *Journal of Market Access Health Policy*. 2015 Mar 12;3.

62.Rothwell PM. Factors that can affect the external validity of randomised controlled trials. *PLoS clinical trials*. 2006;1.

63. Sacristán JA, Aguarón A, Avendaño-Solá C, Garrido P, Carrión J, Gutiérrez A, Kroes R, Flores A. Patient involvement in clinical research: why, when, and how. *Patient Prefer Adherence*. 2016 Apr 27;10:631-40.

64. Tinetti ME, Basch, E. Patients' responsibility to participate in decision making and research. *JAMA*. 2013;309(22): 2331-2332.

65. ZonMw. Een 10 voor patiënten participatie. Online access https://www.zonmw.nl/uploads/tx_vipublicaties/Een_10_voor_patientenparticipatie.pdf. Published 2013. Accessed 31-7-2020, 2020.

66. El Alili M, van Dongen JM, Huirne JAF, van Tulder MW, Bosmans JE. Reporting and Analysis of Trial-Based Cost-Effectiveness Evaluations in Obstetrics and Gynaecology. *Pharmacoeconomics*. 2017;35(10):1007-1033.

67. van Dongen JM, El Alili M, Varga AN, Guevara Morel AE, Jornada Ben A, Khorrami M, van Tulder MW, Bosmans JE. What do national pharmacoeconomic guidelines recommend regarding the statistical analysis of trial-based economic evaluations? *Expert Rev Pharmacoecon Outcomes Res*. 2020;20(1):27-37.

68. Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods. *Stat Med*. 2019 May 20;38(11):2074-2102

10

Summary

SUMMARY

Low back pain (LBP), and sciatica in particular, are highly prevalent conditions that negatively affect a patient's health and quality of life, and are associated with high healthcare and societal costs. Therefore, this thesis aimed to contribute to the development of a sound evidence base on the relationship between LBP, outcomes, and costs *(Chapter 2 and 3)*, the effectiveness and cost-effectiveness of sciatica treatments *(Chapter 4, 5 and 6)*, and to improve scientific methods in LBP research *(Chapter 7 and 8)*.

To address these issues, this thesis was divided into three themes which include:

Theme A: *Relationship between low back pain, outcomes, and costs.*

Theme B: *Effectiveness and cost-effectiveness of sciatica treatments*

Theme C: *Methodological studies*

Below, the results of this thesis will be summarized per theme separately.

Theme A: *Relationship between low back pain, outcomes, and costs.*

Research questions:

1. What is the association between pain severity/disability with health-related quality of life and costs? *(Chapter 2)*
2. Which factors predict high societal costs among chronic low back pain patients? *(Chapter 3)*

Chapter 2 described the longitudinal relationship between pain severity and disability versus health-related quality of life and costs among chronic low back pain patients. A total of 6,316 LBP patients that made up the observational study part of the MINT study were included in the analyses *(Chapter 2)*. The MINT study consisted of 4 randomized controlled trials and an observational study. It was conducted in the Netherlands, with the aim to assess the effectiveness and cost-effectiveness of adding minimal interventional procedures to a standardized treatment program, compared with a standardized treatment

program alone in low back pain patients. The observational study consisted of chronic LBP patients, aged between 18-70 years old, referred to a pain clinic with suspected chronic mechanical LBP and without improvement of symptoms after conservative treatment. Patients who did not want to or were not eligible to participate in the MINT study were invited to participate in the observational study.

This study found pain severity and disability both to have a statistically significant negative longitudinal relationship with health-related quality of life, and a statistically significant positive longitudinal relationship with societal as well as healthcare costs. To illustrate, a clinically relevant increase in disability (defined as a 10 point increase on the 0–100 point ODI) was found to be associated with a decrease in health-related quality of life by 0.096 points (range 0–1), and an increase in societal as well healthcare costs by €170 and €80 per 3-month period, respectively. Based on these results it was concluded that both pain severity and disability are longitudinally related to health-related quality of life, societal costs, and healthcare costs. Disability had a stronger association with all outcomes compared to pain.

Chapter 3 described predictive factors of high societal costs among chronic LBP patients in the Netherlands. Again, 6,316 LBP patients that made the observational study part of the MINT study were included in the analyses *(Chapter 3)*. Having high societal costs (yes/no) was the outcome of this study. High societal costs were defined as the top 10% of cost outcomes. Sensitivity analyses were performed using patients in the top 5% and 20% of cost outcomes. Societal costs were collected using 3-month retrospective cost questionnaires. Prediction models were constructed using backwards logistic regression models. High functional disability, poor physical health, low health-related quality of life, high impact of pain experience, non-Dutch nationality and decreasing pain were found to be predictive of high societal costs in all models, and were therefore considered robust predictors of high societal costs among chronic LBP patients.

Theme B: *Effectiveness and cost-effectiveness of sciatica treatments*

Research questions:

1. Is combination therapy (MDT &TESIs) effective and cost-effective compared to usual care among sciatica patients with an indication for surgery? *(Chapter 4 and 5)*
2. Is exercise therapy in the treatment of sciatica effective? *(Chapter 6)*

Chapter 4 presented the design article of the PLUS-study randomized control trial. The aim of the RCT was to evaluate the effectiveness and cost-effectiveness of combination therapy, consisting of Mechanical Diagnosis and Treatment and Transforaminal Epidural Steroid Injections, compared to no intervention (i.e. usual care) while being on the waiting list for lumbar herniated disc surgery. The RCT had a follow-up of one year. Patients were recruited from seven hospitals in the Netherlands. The targeted sample size was 146. Inclusion criteria included patients with a confirmed case of lumbar disc herniation, an indication for surgery, being aged 18 years or older, who had not received an epidural injection on the same level in the previous six months. The primary outcome was the number of patients undergoing lumbar disc surgery during follow-up. Secondary outcomes included back and leg pain intensity (NPRS), physical functioning (RMDQ-23), self-perceived recovery (GPE), and health-related quality of life (EQ-5D-5L and SF12). For the economic evaluation, societal and healthcare costs during follow-up were measured using questionnaires.

Chapter 5 described the preliminary effectiveness and cost-effectiveness analysis of the PLUS-study. The preliminary analysis was conducted because, patient recruitment lacked behind and we wanted to have a first indication regarding the effectiveness and cost-effectiveness of combination therapy versus usual care. For the preliminary analysis, data from 56 patients who had completed a 6-month follow-up were used. Twenty-seven patients were randomly assigned to combination therapy and 29 to usual care. The results showed that 9 out of 27 patients (33%) received surgery in the intervention group

and 24 out of 29 patients (83%) received surgery in the control group. The adjusted odds ratio of receiving surgery in the intervention group was 0.07 (95%CI: 0.02 to 0.35) compared to the control group. There were no statistically significant differences in clinical effects between both groups. Costs were on average lower by €2,878 in the intervention group compared to the control group. For surgery, the ICER was 1,363, meaning that on average €1,363 were saved per surgery prevented in the intervention group compared to the control group. Hence, the results showed that combination therapy for patients on the waiting list for lumbar herniated might be promising in preventing surgeries compared with usual care, and that there were no differences in clinical effects between both groups. More data and a longer follow-up time are required to see what will happen to the results.

Chapter 6 described a systematic review on the effectiveness of exercise therapy in sciatica patients. A comprehensive literature search to identify relevant randomized controlled trials was performed in PUBMED, EMBASE, Physiotherapy Evidence Database (PEDro), CINAHL, and the Cochrane Library, from the inception of the database to May 2019. To ensure that no articles were missed, the references of included articles were reviewed. Nine RCTs were included, out of which three compared exercise to other therapies, two compared exercise to no therapy, one compared exercise to sham therapy, and three where exercise was provided as an adjunct therapy. No studies compared exercise to surgery. The risk of bias assessment showed that the studies included were of poor methodological quality. Primary outcomes included pain, functional status, and global perceived effect at short, intermediate, and long-term follow-up. The results of the systematic review showed that compared to other therapies, exercise had a small short-term effect on functional status [MD -8.4 (95%CI -15.70, -1.10); 0-100 scale]; a large short-term effect on pain compared to no therapy [MD -2.07 (95%CI -3.24, -0.89), 0-10 scale]; and a medium long-term effect on global perceived effect when given as an adjunct therapy [RR 1.42 (95%CI 1.11, 1.81); RD 0.23; NNT=4]. Other comparisons showed non-statistical significant differences. For all outcomes, the certainty of evidence was low to very low and there was no

explanation for statistical heterogeneity. No one study compared exercises to surgery. Hence, the results of this systematic review showed that the effectiveness of exercise therapy for the treatment of sciatica remains unclear despite its common use in clinical practice. Better studies, in the form of larger, low risk of bias RCTs are highly recommended.

Theme C: *Methodological studies*

Research questions:

1. To what extend is the association between GPE and change in pain and functional status influenced by current health status? *(Chapter 7)*
2. Does the statistical approach in trial-based economic evaluations matter? *(Chapter 8)*

Chapter 7 investigated the construct validity of the Global Perceived Effect (GPE) scale for measuring self-perceived recovery in patients with sciatica. That is, it assessed whether the GPE really measures change in pain and function in sciatica patients over time. Information from 169 postoperative sciatica patients were used. The results showed that GPE was statistically significantly associated with change in leg pain (OR:1.04;95%CI:1.02-1.05), change in back pain (OR:1.02;95%CI:1.01-1.04), and change in functional status (OR:1.08;95%CI:1.04-1.12). However, when current pain and functional status were added to models, the size of some of the associations decreased and the models' explained variance increased. This showed that a patient's current health status influences whether they consider themselves recovered or not on the GPE scale. That is, in judging themselves as recovered or not, patients mainly look at their current health status. In addition, when it was explored whether time duration influences patient scorings, the results showed that time duration did not influence the associations. This indicated that GPE scales do not perform adequately as a transition scale, regardless of transition period. Therefore, in light of these results, the conclusions of this study were that GPE is not a true transition scale for patients with sciatica and that current health status considerably impacts GPE scores, irrespective of recall period.

Chapter 8 explored whether or not accounting for baseline imbalances, skewed costs, correlated costs and effects, and missing data in trial-based economic evaluations has an impact on the results. To accomplish this, data from two trial-based economic evaluations were used. A total of 14 full trial-based economic evaluations were performed per study, in which all of the aforementioned statistical challenges were taken into account step-by-step. Statistical approaches were compared in terms of the resulting cost and effect differences, ICERs, and probabilities of cost-effectiveness. The results showed that, the ICER ranged from 636,744€/QALY and 90,989€/QALY when ignoring all statistical challenges to -7,502€/QALY and 46,592€/QALY when accounting for all statistical challenges, respectively. The probabilities of the intervention being cost-effective at 0€/QALY gained were 0.67 and 0.59 when ignoring all statistical challenges, and 0.54 and 0.27 when all of the statistical challenges were taken into account for the REALISE study and HypoAware study, respectively. To conclude, the study showed that not taking into account the statistical challenges mentioned above may significantly impact the results of trial-based economic evaluation. Therefore, it is of utmost importance to first check the data and to identify statistical challenges that need to be adequately accounted for in the analysis of trial-based economic evaluation data.

SAMENVATTING

Lage rugpijn, en het lumbosacraal radiculair syndroom in het bijzonder, zijn veel voorkomende aandoeningen die een negatieve invloed hebben op zowel de gezondheid als de kwaliteit van het leven van een patiënt. Daarnaast zijn beiden ook geassocieerd met hoge maatschappelijke en gezondheidszorgkosten. Dit proefschrift beoogt daarom bij te dragen aan de ontwikkeling van een solide wetenschappelijke onderbouwing van de relatie tussen lage rugpijn, behandeluitkomsten en kosten *(Hoofdstuk 2 en 3)*, de effectiviteit en kosteneffectiviteit van behandelingen van het lumbosacraal radiculair syndroom *(Hoofdstuk 4, 5 en 6)*, en een verbetering van de gebruikte wetenschappelijke methoden in het onderzoek naar lage rugpijn *(Hoofdstuk 7 en 8)*.

Om deze doelen te bereiken, is dit proefschrift onderverdeeld in drie thema's:

Thema A: *Relatie tussen lage rugpijn, behandeluitkomsten en kosten*

Thema B: *Effectiviteit en kosteneffectiviteit van behandelingen van het lumbosacraal radiculair syndroom*

Thema C: *Methodologische studies*

Hieronder worden de resultaten van dit proefschrift per thema apart samengevat.

Thema A: *Relatie tussen lage rugpijn, behandeluitkomsten en kosten.*

Onderzoeksvragen:

1. Wat is het verband tussen de ernst van de pijn en het functionerenvan de patiënt en zijn of haar gezondheidsgerelateerdekwaliteit van leven en kosten? *(Hoofdstuk 2)*
2. Welke factoren voorspellen het hebben van hoge maatschappelijke kosten bij patiënten met chronische lage rugpijn? *(Hoofdstuk 3)*

Hoofdstuk 2 onderzocht de longitudinale relatie tussen de ernst van de pijn/ functioneren en gezondheidsgerelateerde

kwaliteit van leven en kosten bij patiënten met chronische lage rugpijn. In totaal werden 6.316 chronische lage rugpijnpatiënten die deel uitmaakten van de observationele studie van de MINT-studie meegenomen in de analyses. De MINT-studie bestond in totaal uit 4 gerandomiseerde gecontroleerde studies en één observationele studie. De studie werd in Nederland uitgevoerd met als doel te onderzoeken of radiofrequente denervatie toegevoegd aan een gestandaardiseerd beweegprogramma effectief en kosteneffectief is in vergelijking met een gestandaardiseerd beweegprogramma alleen voor patiënten met chronische lage rugklachten. De observationele studie bestond uit chronische lage rugpijnpatiënten, in de leeftijd tussen 18-70 jaar oud, verwezen naar een pijnkliniek op verdenking van chronische mechanische lage rugpijn en zonder verbetering van de symptomen na conservatieve behandeling. Patiënten die niet wilden deelnemen, of niet in aanmerking kwamen voor deelname aan de MINT-studie werden gevraagd om deel te nemen aan de observationele studie.

Uit deze studie bleek dat ernst van de pijn en functioneren zowel een statistisch significante negatieve longitudinale associate hadden met gezondheidsgerelateerde kwaliteit van leven als een statistisch significante positieve longitudinale associatie met maatschappelijke- en gezondheidszorgkosten. Ter illustratie: een klinisch relevante toename in functioneren (gedefinieerd als een toename van 10 punten op de ODI-schaal van 0-100 punten) was geassocieerd met een afname van de gezondheidsgerelateerde kwaliteit van leven van 0,096 punten (op een schaal van 0-1), en een 3-maandelijkste stijging van zowel de maatschappelijke- als de gezondheidszorgkosten van respectievelijk € 170, - en € 80, -. Op basis van deze resultaten hebben wij geconcludeerd dat zowel ernst van pijn als functioneren een longitudinaal verband heeft met gezondheidsgerelateerde kwaliteit van leven, maatschappelijke kosten en gezondheidszorgkosten. Daarnaast viel op dat deze associatie voor alle uitkomsten het sterkst was voor functioneren.

Hoofdstuk 3 onderzocht welke factoren voorspellend zijn voor hoge maatschappelijke kosten bij chronische lage rugpijnpatiënten in Nederland. Opnieuw werd data gebruikt van 6.316 chronische lage rugpijnpatiënten uit het observationele

cohort van de MINT-studie. Hoge maatschappelijke (ja/nee) was de afhankelijke variabele in deze studie en werd gedefinieerd als het behoren tot de top 10% van de patiënten met de hoogste kosten. Sensitiviteitsanalyses werden uitgevoerd waarbij hoge maatschappelijke kosten gedefinieerd werd als 1) behorende tot de top 5% en 2) behorende tot de top 20%. Data met betrekking tot de maatschappelijke kosten van de patiënten werd verzameld met behulp van driemaandelijkse retrospectieve kostenvragenlijsten. Predictiemodellen werden ontwikkeld met behulp van "backwards" logistische regressies. Functionele beperkingen, een slechte lichamelijke gezondheid, een lage gezondheidsgerelateerde kwaliteit van leven, een hoge impact van pijnervaring, een niet-Nederlandse nationaliteit en een lagere pijnscore bleken in alle modellen voorspellend te zijn voor hoge maatschappelijke kosten, en werden daarom beschouwd als robuuste voorspellers van hoge maatschappelijke kosten bij chronische lage rugpijnpatiënten.

Thema B: *Effectiviteit en kosteneffectiviteit van lumbosacraal radiculair syndroom*

Onderzoeksvragen:

1. Is combinatietherapie (MDT & TESI's) effectief en kosteneffectief in vergelijking met gebruikelijke zorg bij patiënten die een indicatie hebben voor een lumbale herniaoperatie? *(Hoofdstuk 4 en 5)*
2. Is oefentherapie effectief bij de behandeling van het lumbosacraal radiculair syndroom? *(Hoofdstuk 6)*

Hoofdstuk 4 beschrijft de opzet van een gerandomiseerde gecontroleerde studie, de PLUS-studie. Het doel van de PLUS-studie was het evalueren van de effectiviteit en kosteneffectiviteit van een combinatietherapie bij patiënten die een indicatie hebben voor een lumbale herniaoperatie in vergelijking met gebruikelijke zorg (i.e. het niet ontvangen van zorg) terwijl de patiënten op de wachtlijst staan voor een lumbale herniaoperatie. De combinatietherapie bestaat uit mechanische diagnose en behandeling en transforaminale epidurale steroïde-injecties. De studie had een follow-up duur van één jaar. Patiënten werden gerekruteerd uit zeven ziekenhuizen in Nederland. Het

beoogde aantal deelnemers was 146. Inclusiecriteria waren onder meer het hebben van een MRI-bevestigde lumbale hernia, een indicatie voor een lumbale herniaoperatie, 18 jaar of ouder zijn en het niet hebben ontvangen van een epidurale injectie op hetzelfde niveau in de afgelopen zes maanden. De primaire uitkomstmaat was het aantal patiënten dat uiteindelijk een lumbale herniaoperatie ondergaat gedurende follow-up. Secundaire uitkomsten waren onder meer pijnintensiteit in de rug en benen (NPRS), fysiek functioneren (RMDQ-23), zelf ervaren herstel (GPE) en gezondheidsgerelateerde kwaliteit van leven (EQ-5D-5L en SF-12). Voor de economische evaluatie werden de maatschappelijke- en gezondheidszorgkosten van de deelnemers in kaart gebracht. Daar werd onder andere data voor verzameld middels online kostenvragenlijsten bij aanvang van de studie, en 2, 4, 6, 9 en 12 maanden follow-up.

Hoofdstuk 5 beschrijft de voorlopige, tussentijdse, effectiviteits- en kosteneffectiviteitsresultaten van de PLUS-studie. Deze voorlopige analyse werd uitgevoerd omdat de rekrutering van patiënten achterbleef, maar we wel een eerste indicatie wilden hebben met betrekking tot de effectiviteit en kosteneffectiviteit van de combinatietherapie in vergelijking met gebruikelijke zorg. Voor de voorlopige analyse werden gegevens van 56 patiënten gebruikt die een follow-up van 6 maanden hadden voltooid. Zevenentwintig patiënten werden op basis van kans (randomisatie) toegewezen aan de combinatietherapiegroep (interventiegroep) en 29 ontvingen gebruikelijke zorg (controle groep). De resultaten lieten zien dat na 6 maanden, 9 van de 27 interventiegroep patiënten (33%) een lumbale herniaoperatie hadden ondergaan, terwijl dit in de controlegroep het geval was voor 24 van de 29 patiënten (83%). De bijbehorende “odds ratio” voor het ondergaan van een lumbale herniaoperatie in de interventiegroep was 0,07 (95% BI: 0,02 tot 0,35) in vergelijking met de controlegroep. Er waren geen statistisch significante verschillen tussen beide groepen in de klinische effecten. Daarnaast waren de maatschappelijke kosten in de interventiegroep gemiddeld genomen € 2.878 lager dan in de controlegroep. Voor het ondergaan van een lumbale herniaoperatie (ja/nee) was de ICER 1.363 Dit betekent dat er gemiddeld genomen € 1.363

bespaard werd per vermeden operatie in de interventiegroep in vergelijking met de controlegroep. De resultaten toonden dus aan dat combinatietherapie voor patiënten op de wachtlijst voor een lumbale hernia een veelbelovende behandeling kan zijn om lumbale hernia operatie te voorkomen. Dit resultaat is echter nog onzeker, omdat de onderzoekspopulatie nog redelijk klein is en we de data behorende bij 12 maanden follow-up nog niet hebben geanalyseerd.

Hoofdstuk 6 beschrijft een systematische review naar de effectiviteit van oefentherapie bij lumbosacraal radiculair syndroom patiënten. We hebben een uitgebreide literatuurzoekopdracht uitgevoerd om relevante gerandomiseerde gecontroleerde studies te identificeren in PUBMED, EMBASE, Physiotherapy Evidence Database (PEDro), CINAHL en de Cochrane Library, welke gepubliceerd waren vanaf publicatie van de database tot mei 2019. Om ervoor te zorgen dat er geen artikelen werden gemist, hebben we daarnaast de referentielijsten van de geïncludeerde studies nagelopen. In totaal werden negen studies geïncludeerd, waarvan er drie oefentherapie vergeleken met andere therapieën, twee oefentherapie vergeleken met "geen behandeling", één oefentherapie vergeleken had met een placebo en drie oefentherapie had onderzocht als aanvulling op andere vormen van therapie. Geen enkele studie vergeleek oefentherapie met chirurgie. De "risk of bias assessment" liet zien dat de geïncludeerde studies een relatief lage methodologische kwaliteit hadden. De primaire uitkomstmaten van de geïncludeerde studies waren pijn, functioneren en ervaren herstel na een korte, middellange en lange follow-up duur. De resultaten van de systematische review lieten zien dat oefentherapie in vergelijking met andere therapieën op korte termijn een klein effect had op functioneren [MD -8,4 (95% BI -15,70, -1,10); Schaal 0-100]; in vergelijking met "geen therapie" op korte termijn een groot effect had op pijnintensiteit [MD -2,07 (95% BI -3,24, -0,89), 0-10 schaal]; en als aanvullende therapie een "medium" effect had op ervaren herstel op de lange termijn [RR 1,42 (95% BI 1,11, 1,81); RD 0.23; NNT = 4]. Andere vergelijkingen lieten geen statistisch significante verschillen zien. Voor alle uitkomsten was er sprake van een hoge mate van onzekerheid en er was geen

verklaring voor de gevonden statistische heterogeniteit. Op basis van deze resultaten concludeerden wij dat de effectiviteit van oefentherapie als behandeling voor lumbosacraal radiculair syndroom onduidelijk is, ondanks dat deze in de praktijk veel gegeven wordt. Betere studies, in de vorm van grotere gerandomiseerde gecontroleerde studies met een lagere "risk of bias" worden daarom sterk aanbevolen.

Thema C: *Methodologische studies*

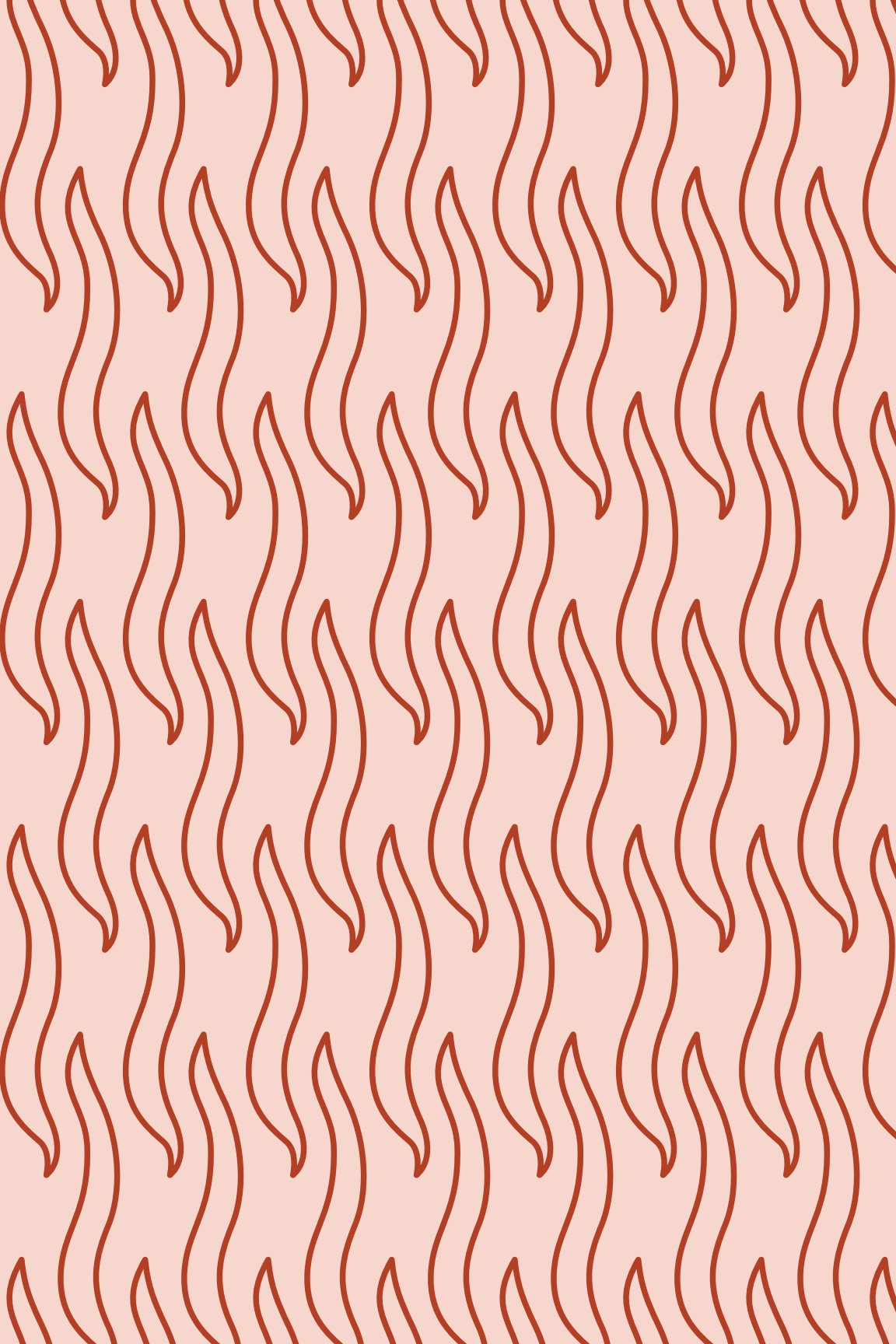
Onderzoeksvragen:

1. In hoeverre wordt het verband tussen ervaren herstel, gemeten met de Global Perceived Effect GPE, en verandering in pijn en functioneren beïnvloed door de huidige gezondheidstoestand van de patiënt? *(Hoofdstuk 7)*
2. In hoeverre doet de gekozen statistische model ertoe in een economische evaluatie die uitgevoerd wordt op basis van trial data? *(Hoofdstuk 8)*

Hoofdstuk 7 onderzocht de constructvaliditeit van de Global Perceived Effect (GPE), een schaal die veel gebruikt wordt voor het meten van ervaren herstel bij lumbosacraal radiculair syndroom patiënten. Met andere woorden; meet de GPE-schaal daadwerkelijk veranderingen in pijn en functioneren? Om deze vraag te beantwoorden is gebruik gemaakt van data van 169 postoperatieve lumbosacraal radiculair syndroom patiënten. De resultaten toonden aan dat de GPE-schaal statistisch significant geassocieerd was met veranderingen in beenpijn (OR: 1,04; 95% BI: 1,02-1,05), rugpijn (OR: 1,02; 95% BI: 1,01-1,04) en functioneren (OR: 1,08; 95% -BI: 1,04-1,12). Echter, toen de huidige pijn en functionele status scores (de scores dus op het moment van afname) van de patiënten werden toegevoegd aan de modellen, nam de omvang van de associaties af en de verklaarde variantie van de modellen toe. Dit suggereert dat de gezondheidstoestand op het moment van afname van een patiënt grote invloed heeft op of hij of zij zich als "hersteld" scoort op de GPE-schaal. Dat wil zeggen, bij het beoordelen van ervaren herstel kijken patiënten vooral naar hun huidige gezondheidstoestand en niet naar de verbetering of verslechtering daarvan. Bovendien, bleek dat de tijdsduur geen invloed had op deze resultaten. Op basis van deze

bevindingen hebben wij geconcludeerd dat de GPE-schaal geen echte ”transition scale” (veranderschaal) is.

Hoofdstuk 8 onderzocht de impact van het al dan niet corrigeren voor eventuele baseline verschillen, de doorgaans scheve verdeling van kostendata, de doorgaans aanwezige correlatie tussen kosten en effecten, en de doorgaans aanwezige missende data in economische evaluaties. Hiervoor werd gebruik gemaakt van gegevens van twee reeds gepubliceerde economische evaluaties. Per studie werden in totaal 14 volledige economische evaluaties uitgevoerd, waarbij stap voor stap rekening werd gehouden met alle bovengenoemde statistische issues. De diverse statistische modellen werden vergeleken in termen van de geschatte verschillen in kosten en effecten, ICER’s en kansen op kosteneffectiviteit. De resultaten lieten zien dat de ICER varieerde van 636.744 €/QALY (REALISE studie) en 90.989 €/QALY (HypoAware studie) wanneer alle bovengenoemde issues werden genegeerd, tot -7.502 €/QALY (REALISE studie) en 46.592 €/QALY (HypoAware studie) wanneer met alle issues rekening werd gehouden. Daarnaast verschilde de kans op kosteneffectiviteit bij een betalingsbereidheid van 0 €/QALY van 0,67 (REALISE studie) en 0,59 (HypoAware studie) wanneer alle issues werden genegeerd tot 0,54 (REALISE studie) en 0,27 (HypoAware studie) wanneer met alle issues rekening werd gehouden. Op basis van deze resultaten concludeerden wij dat het gebruikte statistische model een aanzienlijke invloed kan hebben op de resultaten van een economische evaluatie. Daarom is het van het grootste belang om eerst de in een economische evaluatie een statistische strategie te kiezen die zo goed mogelijk aansluit bij de verzamelde data.



About the author

11

About the author

ABOUT THE AUTHOR

Elizabeth Nyasha Mutubuki was born on the 4th of December 1983 in Harare, Zimbabwe. She completed her GCSE and A'Level at St Dominic's Chishawasha. She holds a Bachelor's degree in Physiotherapy from the Amsterdam University of Applied Sciences (Hogeschool van Amsterdam). Her interest in public health was stimulated during her internships in physiotherapy which took place in different countries, UK (Peterborough and Stamford Hospital NHS Foundation Trust), Zimbabwe (St Anne's Hospital) and, Netherlands (Elsinga & Meijer Fysiotherapie). In 2010 she received a Master's degree in Public Health with a specialization in Health Economics and Health Promotion from Karolinska Institutet Medical University in Stockholm, Sweden.

Following her Master's degree, Elizabeth worked in diverse physiotherapy practices in the Netherlands. After which Elizabeth, started her own her own physiotherapy practice, Fysio Praktijk Berghuis/Fysiobezorgd where she focused mostly on the treatment and management of patients with musculoskeletal disorders. In 2013 she started Hutano, a Public health consultancy. During her time at Hutano Elizabeth performed together with Koninklijk Instituut voor de Tropen (KIT), a desk literature review of institutional arrangements for health financing in selected African countries. The literature review was done under the ReBuild Programme of health systems research and stakeholder dialogue and capacity building. The Rebuild Programme seeks to move from the immediate recovery measures to longer term measures for Universal Health Coverage (UHC), which encompasses equity in access and coverage, a cause dear to her.

In November 2016, Elizabeth started her PhD position at the Faculty of Health Sciences and department of Health Technology assessment at the Vrije Universiteit Amsterdam. There she worked as junior researcher on the PLUS-study (Preventing Lumbar Disc hernia trial) and conducted a randomized controlled trial in the Netherlands. Next to the trial, Elizabeth conducted a systematic review, prediction model, longitudinal

data analysis, effectiveness and cost-effectiveness analyses and methodological papers. Her research project focused on low back pain; treatments, health outcomes and costs. In addition Elizabeth attended several courses, during this period.

Elizabeth currently holds a post-doctoral position at the department of Epidemiology and Data Science of the Amsterdam Universitair Medische Centra. Her current position focuses on the methodology of guideline development. In addition Elizabeth is a board member of the Friendship Bench, an initiative providing sustainable community based psychological interventions that are evidence based, accessible and scalable in Zimbabwe and beyond. Elizabeth is one of the co-founders of the ZimNetwork, a Zimbabwean Platform promoting professional interaction of Zimbabweans in the Benelux.

Elizabeth lives in Zeist with her husband, Jonathan and 2 daughters Savannah (6 years) and Aurora (3 years).

PHD PORTFOLIO

Name	Elizabeth Nyasha Mutubuki
Affiliation	Vrije Universiteit Amsterdam
PhD period	November 2016 –February 2020
Promotoren	prof.dr. R.W.J.G. Ostelo
	prof.dr. M.W. van Tulder
Co-promotoren	dr. J.M. van Dongen
	dr. C.L.A. Vleggeert-Lankamp

Education	**Year**	**ECTS**
EpidM, VUmc, Amsterdam		
Epidemiological research: design and interpretation (V10) Epidm	2017	4
Missing data: consequences and solutions (WK81) Epidm VUmc, Amsterdam	2019	2
Vrije Universiteit, Amsterdam		
Economic Evaluation VU	2018	6
Advanced statistics VU	2018	6
Career orientation	2019	1
VUMC Academie		
Research integrity	2017	2
BROK	2018	1.5
McKenzie		
McKenzie therapie	2017	1.5
	Subtotal:	*24*
Congresses		
International		
International forum for back and neck pain research, Oslo, Norway	2017	2
International forum for back and neck pain research, Quebec Canada	2019	2
ISPOR, Copenhagen Denmark	2019	2
National		
AMS, Amsterdam, the Netherlands	2017	2
KNGF, Barneveld, the Netherlands (oral presentation)	2017	1.5
	Subtotal:	*9.5*
	Total:	**33.5 ECTS**

LIST OF PUBLICATIONS

This thesis

Mutubuki EN, van Helvoirt H, van Dongen JM, Vleggeert-Lankamp CLA, Huygen FJPM, van Tulder MW, Klopper-Kes HAHJ, Ostelo RWJG. Cost-effectiveness of combination therapy (Mechanical Diagnosis and Treatment and Transforaminal Epidural Steroid Injections) among patients with an indication for a Lumbar Herniated Disc surgery: Protocol of a randomized controlled trial. *Physiother Res Int.* 2020 Jan;25(1):e1796.

Mutubuki EN, Beljon Y, Maas ET, Huygen FJPM, Ostelo RWJG, van Tulder MW, van Dongen JM. The longitudinal relationships between pain severity and disability versus health-related quality of life and costs among chronic low back pain patients. *Qual Life Res.* 2020 Jan;29(1):275-287. doi: 10.1007/s11136-019-02302-w. Epub 2019 Sep 17.

Mutubuki EN, Luitjens MA, Maas ET, Huygen FJPM, Ostelo RWJG, van Tulder MW, van Dongen JM. Predictive factors of high societal costs among chronic low back pain patients. *Eur J Pain.* 2020 Feb;24(2):325-337.

Mutubuki EN, van Helvoirt H, van Dongen JM, Ben Jornada A, Vleggeert-Lankamp CLA, Huygen F, van Tulder MW, KlopperKes AHJ, Rooker S, Leanders M, van der Gaag NA, Hoffmann CFE, Leilifeld P, Ostelo RWJG, on behalf of the PLUS Study Research Group. Effectiveness and cost-effectiveness of combination therapy (Mechanical diagnosis & treatment & Transforaminal Epidural Steroid Injections) for patients with an indication for a lumbar herniated disc surgery: A randomized controlled trial. (to be submitted)

Seiger A, **Mutubuki E.N**, van Tulder M.W, VleggeertLankamp C.L.A.M, Peul W.C, van Susante J.L.C., Harhangi B.S., Ostelo R.W.J.G, Jansma E.P., Rubinstein S.M. Exercise therapy for sciatica. Is it effective? A systematic review and meta-analysis. *(under review)*

Elizabeth N Mutubuki, Heleen Kuiper, Johanna M van Dongen, Maurits W van Tulder, Steven J Kamper, Carmen LA Vleggeert-Lankamp, Teddy Oosterhuis, Raymond WJG Ostelo. Current health status strongly influences the Global Perceived Effect scale among sciatica patients who received lumbar disc surgery. *(under review)*

***Elizabeth N. Mutubuki**, *Mohamed El Alili, Judith E. Bosmans, Teddy Oosterhuis, Frank Snoek, Raymond W. J. G. Ostelo,Maurits W. van Tulder, Johanna M. van Dongen. The statistical approach in trial-based economic evaluations matters: get your statistics together! *(under review)*
**Contributed equally*

Other publications

Vaughan K, Toonen J, **Berghuis-Mutubuki E**, Dieleman M (2014) Desk Review of institutional arrangements for health financing in selected African countries, KIT Netherlands with TARSC, MoHCC Zimbabwe

Mutubuki, E., El Alili, M., Oosterhuis, T., Bosmans, J., Ostelo, R. W., van Tulder, M. W. & van Dongen, H., A comparison of statistical methods used in trial-based economic evaluations; Does it matter which method is used? Nov 2019, In : *Value in Health*. 22, p. S705-S705.

Elizabeth N. Mutubuki. Predictive factors of high societal costs among chronic low back pain patients-wetenschap summary. Koninklijk Nederlands Genootschap voor Fysiotherapie, *FysioPraxis -Vakblad voor de fysiotherapeut*. June 2020, p. 31.

ACKNOWLEDGEMENTS

The end is nigh! ☺

"There is a tide in the affairs of men
Which, taken at the flood, leads on to fortune;
Omitted, all the voyage of their life
Is bound in shallows and in miseries.
On such a full sea are we now afloat;
And we must take the current when it serves,
Or lose our ventures."

William Shakespeare

In November 2016 I took the current when it served and started this PhD journey. Today as I sit to reflect on my journey for the past years and acknowledge those that made it profound, I am made aware that the years too in themselves, were in many ways profound. The beginning was marked by an end to "yes we can" threatening the possibility to affordable healthcare, a cause dear to me. A cause that I hope this thesis in many ways contributes to. Today as I put pen to paper, we sit in a different kind of reality, in which our norms are tested, in which we are reminded that democracy is a very fragile state, whose success depends on us to uphold it. Just like science, it is our responsibility to protect and guard its integrity by respecting the results/outcomes.

The period saw the realization of the PLUS-study trial, which required of me a new kind of resilience. I assumed new identities, that of author in the academic world, PhD of Raymond and Maurits, and that of a mother of two. Aurora dawned on us, reminding us to keep looking up! Savannah learnt to read. I learnt to dress my PowerPoint presentations in academic regalia! ☺ New collaborations were formed and new knowledge was acquired. Different perspectives were gained into the world of neurosurgery, and Henry Mash's Do No Harm book broadened them further. A new understanding of the meshwork that knits the structures of our healthcare was attained, one of the privileges of running a randomized controlled trial, I guess.

However let us always remember that, running a randomized controlled trial is exceedingly onerous just like a PhD. Hence I am grateful for the unwavering support of my family, friends and colleagues during my PhD. This section of the thesis is dedicated to you. I am deeply indebted to your love and kindness.

"If I have seen further it is by standing on the shoulders of Giants."

Isaac Newton

My supervisors, the academic giants whose shoulders, I have stood on to see further into the academia, thank you for nurturing this talent and being a source of inspiration.

Raymond: It has been and continues to be a great pleasure to work with you. The enthusiasm with which you teach and impart knowledge is nothing short of admirable. Over the years I have watched with great appreciation how you manage to bring together conflicting ideas or sides. I appreciated that you had an eagle eye, that did not miss any inconsistencies in a paper or the practical application of the results, leaving behind an article close to perfection! Thank you the wisdom gained and acknowledging my abilities!

Hanneke, like I said before "every PhD needs a Hanneke"! ☺ Thank you for being a cool co-promotor and always having a strategy out of any situation. ☺ I enjoyed the quick lunch trips to sample sales and the book exchanges. Glad that you made yourself available, to answer questions (big or small), to go back over an analysis or to scream loudly when Michelle Obama entered the room. Our meetings were very comprehensive, half work and half work related issues. ☺ Thank you for not holding me hostage to stereotypes, ☺ and of course for imparting that statistical wizardry!

Maurits: Your ability to maintain calm even when an RCT was grinding to a halt, is beyond amazing! Thank you for the wise counsel and for believing in my abilities. You redefine leadership. It was really great working with you.

Carmen: Thank you for recruiting patients for the PLUS-study and always looking for ways to improve the study.

Leden van de leescommissie: prof.dr. J.E. Bosmans, prof. dr. W.C. Peul, prof.dr. C. Veenhof, dr. G.A. de Wit, dr. J.L.C. van Susante, *hartelijk dank voor jullie aanwezigheid en voor de tijd en energie die jullie hebben gestoken in het lezen en beoordelen van mijn proefschrift. Ik kijk uit naar jullie vragen tijdens mijn verdediging en een mooie afronding.*

I would also like to extend my gratitude to **my co-authors**, Y. Beljon, E.T. Maas, F.J.P.M. Huygen, R.W.J.G. Ostelo, M.W. van Tulder, J.M. van Dongen, M.A. Luitjens, H. van Helvoirt, C.L.A. Vleggeert-Lankamp, A.H.J. Klopper-Kes, A. Ben Jornada, S. Rooker, M. Leanders, N.A. van der Gaag, C.F.E. Hoffmann, P. Leilifeld, A. Seiger, J.L.C. van Susant, B.S. Harhangi, E.P. Jansma, S.M. Rubinstein, H.J. Kuiper, S.J. Kamper, T. Oosterhuis, M. El Alili, J.E. Bosmans, F. Snoek. Thank you for the feedback on the articles that helped improve their quality.

De PLUS-studie trial is *uitgevoerd in samenwerking met medewerkers van Hagaziekenhuis, Spaarne Gasthuis, Medische Spectrum Twente, Reinier de Graaf, Kliniek ViaSana, Amphia Ziekenhuis, Caniscius Wilhemina Ziekenhuis en de Rugpoli. In het bijzonder wil ik ook Syta, Yvette en Connie even iets uitgebreider noemen.* ***Syta****, bedankt dat je altijd op zoek was naar patiënten voor de PLUS-studie.* ***Yvette****, bedankt voor je enthousiasme voor de PLUS-studie en je ideeën om de patiënten inclusie te verbeteren.* ***Connie****, wat fijn dat je bereid was om je bureau met mij te delen toen we patiënten gingen bellen voor de PLUS-studie. Verder wil ik ook* ***Dr. Rooker*** *bedanken voor zijn doorzettingsvermogen voor het includeren van patiënten voor de PLUS-studie.*

"Education is the most powerful weapon which you can use to change the world."

Nelson Mandela

My gratitude also extends to my colleagues at the VU department of Health Sciences, whose academic journeys crossed with

mine. I enjoyed the chats in the corridors, the friendly smiles, the drinks, exchanges, the dinners, trips and collaborations. In particular I would also like to acknowledge **Florine**. Our journeys started about the same time. I will never forget our trips to the METC office for clarity with setting up our randomized controlled trials. I am glad that we could share our experiences. All the best as you finalize your last chapters. I will be rooting for you! To my officemates, **Alessandro**, **Allan**, **Jesper** and **Mosdje** and at one point **Gisela**. Thanks you for making the office fun. **Mosdje** thank you for the updates you brought about to the office and creating a gezellig atmosphere. **Allan**, thank you for your contagious humor and jokes that kept us cracking hours on end. My last MSK meeting was amazing, thank you for that lovely send off. Wishing you well as you too conclude your PhD and embark on a new mission. **Alessandro** thank you for the tips and introductions at the international conferences. It made navigating these spaces easier. I am glad that I could draw from your experiences and that of Jesper. **Jesper**, thank you for the pointer and chats. I enjoyed sharing the office with you. To the **U-4 wing**, **Angela, Mo**, **Ale**, **Anita**, **Agi**. and **Judith**, thank you for the wonderful time at the ISpor conference and international lunches shared. Mo, methodology dude, ☺ I enjoyed and learnt a from you, when we worked together our notorious methodological paper. **Prof. Bosmans** (not incoming anymore ☺), dear Judith. I am in awe of how you managed to create a lovely HTA team. A group of intellectuals but also a nice little "family", that I have become part of. ☺

" Zvamunoona husahwira hunokunda hukama"

Oliver Tuku Mtukudzi

To all my wonderful friends, mentioned and unmentioned here, thank you for your support and understanding during this period. Looking forward to catching up without PhD interference. ☺ **Linda**, that all-nighter you pulled reading my thesis and even coming up with a question was epic! Thank you for being the air beneath my wings, *hanzi professor vedu*. **Doc**, the VU will finally officiate a title that we assumed many moons ago in protest against the academic heist of our time. Thank you for your

unequivocal friendship and all that champagne to celebrate my publications. **Karen**, I remember a huge bouquet of flowers when I got accepted on this PhD. Thank you for your support and confidence in my abilities. **Shingi**, asahwira hatikutyirei imimi! Chero Offy said you got this. Thank you for believing in me, and for sure we got this, because you got me. ☺ **Rashna** how nice that you always came up with great ideas to do with the kids and was available for a chat. **Miriam**, I am glad that we also could find time for a hearty cackle even in less happy PhD times. **Mary**, I remember that one of my many PhD interviews was at your house on skype, in between patient treatments. Thank you for the support the yummy plates of *sadza nenyama nemuriwo*, that always managed to soothe the soul after a day of trying to include patients in the trial. Thank you **Thandi**, **Maria** and **Anna**. I am glad that we could share our PhD journeys. To my **Kule B**, *sahwira mukuru* with whom I deeply share the love of Tuku, *rudo moto moto*! **Bella**, **Nono**, **Henriette**, **Lena**, **Zakile**, **Patie**, **Michelle**, **Merel**, **Farai**, **Nyasha**, thank you for your friendship and support. My gratitude also goes to my many **book clubs** that provided a nice diversion during this PhD.

"All happy families are alike; each unhappy family is unhappy in its own way."

Leo Tolstoy

To my **families**, **immediate**, **extended** and **inlaws**, thank you for the nice holidays, keeping the kids entertained and for rooting for me. To my **ambuya** and **sekuru** who followed me to Quebec thank you for believing in me. You are the definition of amazing! To all my nieces, nephews, godchildren, little boys and girls that could not see themselves in the academia. May you see yourselves. To **Dylan** without whom I would not have had the good fortune of being called tete, this book is for you too. To my brothers, **Tate** and **Kudzi** I love you dearly and without reason! Thank you for reminding me to enjoy myself during my PhD and good luck with keeping me calm for the defense! ☺

"Learn everything. Fill your mind with knowledge—it's the only kind of power no one can take away from you."

Min Jin Lee

To **my parents**, my heroes, this one is especially for you, *vabereki vadhokotera*! ☺ Thank you for everything, all sacrifices made, for believing in me and encouraging me to learn. Thank you for being there when I need you and for taking pride in what I do. You brighten up my days. Looking forward to celebrating this one with you.

"I matter. I matter equally. Not "if only." Not "as long as." I matter equally. Full stop."

Chimamanda Ngozi Adichie

To my two most precious, **Aurora** and **Savannah**, you fill my heart with happiness! This book is for you, let it remind you to believe in yourself, to be resilience, that you matter equally and most importantly that you are loved beyond comprehension!

"My baby don't care for shows
My baby don't care for clothes
My baby just cares for me"

Nina Simone

To my love, **Jonathan**, what a journey it has been. Thank you for your love and support, for tears shed, the champagne celebrations for my achievements along the journey and for holding the fort when I was away for conferences, courses or when I just needed to get my writing done. Love lives here! ♡ Looking forward to resuming our LP evenings, Nina Simone has gathered enough dust!

Finally if I could borrow from Savannah, may this add colour to the academic rainbow!

O, fly and never tire,
Fly and never tire,
Fly and never tire,
There's a great camp-meeting in the Promised Land.

— From an African American spiritual

NOTES

Tatenda!

The relationship between pain, disability, health-related quality of life and costs over time is explored. Disability resulted in higher societal costs compared to pain. Predictors of high societal costs are identified. Further the effectiveness and cost-effectiveness of sciatica treatments is investigated. The construct validity of the GPE is explored as well as the impact on results of correcting for statistical challenges in trial based economic evaluations.

Elizabeth Nyasha Mutubuki

Amsterdam Movement Sciences

VU Vrije Universiteit Amsterdam

Amsterdam UMC
University Medical Centers

Amsterdam Movement Sciences conducts scientific research to optimize physical performance in health and disease based on a fundamental understanding of human movement in order to contribute to the fulfillment of a meaningful life.

Koninklijk Nederlands Genootschap voor Fysiotherapie