Randomised controlled trials reflect clinical practice when comparing the course of low back pain symptoms in similar populations

Esther T Maas[1,2§]; Johanna M van Dongen[1,2]; Johan NS Juch[3]; J George Groeneweg[3]; Jan-Willem Kallewaard[4]; Michiel R de Boer[1,2], Bart Koes[5]; Arianne P Verhagen[5,6]; Frank JPM Huygen[3]; Maurits W van Tulder[1,2,7]; Raymond WJG Ostelo[1,2,8]

1. Department of Health Sciences, Faculty of Science, Vrije Universiteit Amsterdam, The Netherlands
2. Amsterdam Movement Science research institute, 1081HV Amsterdam, The Netherlands
3. Department of Anaesthesiology, Erasmus Medical Centre, Rotterdam, The Netherlands
4. Department of Anaesthesiology, Rijnstate hospital, Velp, The Netherlands
5. Department of General Practice, Erasmus Medical Centre, Rotterdam, The Netherlands
6. Graduate School of Health, Discipline of Physiotherapy, University of Technology Sydney, Ultimo, NSW 2007, Australia
7. Department of Physiotherapy and Occupational Therapy, Aarhus University Hospital, Aarhus, Denmark
8. Department of Epidemiology and Biostatistics, Amsterdam UMC, Vrije Universiteit, The Netherlands

§ Corresponding author: Esther T Maas, University of British Columbia, Faculty of Medicine, School of Population and Public health, 2206 East Mall, Vancouver, BC V6T 1Z3, Canada; E: esther.maas@ubc.ca; T: +1-604-8221228

ABSTRACT

OBJECTIVE: This study compares participants in randomized controlled trials (RCTs) (the Minimal Invasive Treatment [MinT] trials) to participants in a related observational study with regard to their low back pain (LBP) symptom course.

STUDY DESIGN AND SETTING: Eligible patients were diagnosed with chronic LBP originating from the facet joints (N = 615) or sacroiliac (SI) joints (N = 533) and were treated with radiofrequency denervation and an exercise program. Randomized patients were compared to patients in the related observational study who fulfilled all RCT eligibility criteria (observational group 1) and to patients who did not fulfill at least one of the RCT eligibility criteria (observational group 2). Outcomes were pain intensity, treatment success, and functional status over a 3-month period. Longitudinal mixed-model analyses and linear regression models were applied to analyze the differences in outcomes between the RCT and observational study groups.

RESULTS: No differences in symptom course were found between patients in the RCTs and patients in observational group 1. Patients with facet joint pain in observational group 2 had overall less treatment success (odds ratios [OR], 0.67; 95% confidence interval [CI], 0.50-0.90), and less improvement in physical functioning (mean difference [MD], 5.82; 95% CI, 2.54-9.11) compared to the RCT patients. Patients with SI joint pain in observational group 2 had higher pain scores (MD, 0.40; 95% CI, 0.09-0.72), less treatment success (OR, 0.72; 95% CI, 0.54-0.96), and less improvement in physical functioning (MD, 7.16; 95% CI, 3.84-10.47) compared to the RCT patients.

CONCLUSION: This supports the generalizability of results from the MinT RCTs as this study suggests that these RCTs reflect clinical practice when comparing similar populations. To what extent this holds true for all RCTs in LBP should be further explored.

What is new?

Key findings

- This study suggests that the randomized controlled trial (RCT) results of the MinT trials are comparable to results of an observational study with a similar intervention and population, and only small differences are shown when the population in the observational study differs from the RCTs.

What this adds to what was known?

- The present study adds to the existing literature that results of pragmatic RCTs in a secondary care setting for patients who received radiofrequency denervation at the pain clinic are comparable to the results of this treatment for similar patients in daily clinical practice.

What is the implication and what should change now?

- Our study suggests that RCTs do reflect clinical practice when comparing similar populations and can increase the generalizability of results. Observational studies can be taken into account when assessing the symptom course after an intervention as long as the clinical features of the intervention and the included population are comparable.

INTRODUCTION

There is on-going debate whether results of randomised controlled trials (RCT) can be extrapolated to patients in routine care setting[1–4]. It has been suggested that the willingness of patients to be randomly allocated to a treatment differentiates these individuals from the average patient, and therefore participation in an RCT might influence the course of symptoms[5]. In addition, the strict eligibility criteria of patients participating in RCTs challenge the generalizability of RCT results[6,7]. Results of well-designed observational cohort studies, in turn, are presumed to resemble daily practice more closely [1–3,8,9]. This raises the question to what extend the outcomes and symptom course in patients included in RCTs and observational studies are comparable.

Evidence generated in observational studies is often ignored in systematic reviews as the assumption is that their findings might be biased[10].One recent meta-analysis compared outcomes of RCTs and observational studies in the field of low back pain (LBP), and showed that the clinical course of LBP symptoms after a treatment in primary care followed a pattern that was similar using both study designs[5]. Difficulties with RCTs have been acknowledged in the assessment of surgical interventions in spinal disorders, and observational studies can be a valuable contribution to the existing knowledge[11]. However, it is unknown how well these comparisons between RCTs and observational studies are transferrable to a population of patients with chronic LBP in a secondary care setting.

The MinT (Minimal Invasive Treatment) study was designed to evaluate the effectiveness and cost-effectiveness of radiofrequency (RF) denervation added to an exercise programme for patients with chronic LBP[12]. RF denervation is a technique that attempts to modulate neural transmission of nociceptive stimuli, reducing spinal pain. It aims to denaturalise the nerves by applying an electric current (heat). This would prevent the conduction of nociceptive impulses[13,14]. RF denervation is a commonly used treatment in patients with LBP originating from the facet joints and sacro-iliac (SI) joints [15,16]. These sources of pain are also named as mechanical LBP[16,17] and are generally assumed to be separate sources. As such, much of the literature distinguishes the entities as did we in a previous publicaiton[12].

The MinT study provides an excellent opportunity to compare results of RCT data with observational study data, because the study consisted of three RCTs and an observational study; and the vast majority of patients in the Netherlands who were treated with RF denervation during the inclusion period

participated in the MinT study. More details on the study design and participants can be found in the study protocol and the publication on effectiveness results[12,17].

The aim of this study was to assess the generalizability of the results from RCTs by comparing the course in LBP symptoms over a 3-month period in (1) randomised study groups, (2) observational study groups with patients that fulfil the eligibility criteria of the RCTs, and (3) patients in clinical practice who do not fulfil at least one of the eligibility criteria for the RCTs. This design allows us to investigate the sole impact of randomisation, as well as differences in results caused by the selection of participants in RCTs.

METHODS

Study design and setting

The MinT study was a Dutch nationwide, multicentre study conducted in 16 pain clinics and 102 physiotherapy practices[12]. The Medical Ethics Committee of the Erasmus University Medical Centre in Rotterdam granted ethical approval (registration number MEC-2012-079). All included patients gave written informed consent.

Participants

The MinT study originally consisted of four RCTs (including patients with 1) facet joint pain, 2) SI joint pain, a combination of symptoms, and 4) discogenic pain) and an observational study. One trial was designed to evaluate RF denervation for pain from the intervertebral disks. This trial was prematurely terminated because of lack of eligible patients and not included in this analysis. Tah data from the RCT for patients with a combination of symptoms was not included either because most patients in the observational study were not identified at baseline with a combination of symptoms. As such these data could not be extracted from the observational study. Patients in the facet joint trial trial were included between January 1, 2013 and June 3, 2014. Patients in the SI joint trial were included between January 1, 2013 and July 1, 2014 (see Table 1).

Observational study group inclusion criteria

Patients were eligible for the MinT study 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. For the analysis of the present study, patients in the observational groups were

included between the date of the inclusion period of the RCTs closed and December 17, 2015 (see Table 1). These patients did not have the choice to be randomised, and hence resemble daily practice.

RCT inclusion criteria

Extra inclusion criteria for patients in the RCTs were: age between 18 and 70 years and having a positive diagnostic facet joint or SI-joint block (≥50% pain reduction 30-90 minutes after procedure). Extra exclusion criteria for the RCTs were pregnancy, anticoagulant drug therapy and/or coagulopathy, body mass index (BMI) >35, involved in a work-related conflict, and severe psychiatric or psychological problems. These eligibility criteria used in the study protocol were standardized and determined by pain physicians to ensure that the study population would be eligible for RF denervation in clinical practice in the Netherlands.

More details on the eligibility criteria are reported in the study protocol[12].

In summary, we compared RCT patients to patients in the observational study; all with chronic LBP originating from the facet joints or SI-joints, receiving RF denervation and an exercise programme:

- Randomised study group: Intervention group of the facet joint or SI-joint RCT, receiving RF denervation and physiotherapy.
- Observational study group 1: Patients who fulfilled all RCT eligibility criteria, but were not randomised, and self-reported to have received RF denervation and physiotherapy.
- Observational study group 2: Patients who did not fulfil at least one of the RCT eligibility criteria, were not randomised, and self-reported to have received RF-denervation and physiotherapy. These could, for example, be patients who received RF denervation but were older than 70 years, or with a BMI >35.

Study interventions

The randomized study group received RF denervation plus a 3-month standardized exercise program combined with psychological support if necessary. RF denervation included facet joint RF denervation, or Cooled RF denervation, Simplicity III probe or Palisade technique as treatment for SI joint pain [12,19,20]. Patients were asked to refrain from any cointervention during the 3-month intervention period. Anesthesiologists at the participating pain clinics recruited the patients and carried out diagnostic blocks and RF denervation. Every participating pain clinic had a referral agreement with physiotherapy practices in their region to provide the standardized exercise program. The psychological interventions, if necessary, took place in a primary care setting.

Patients in the observational study received usual care and were monitored prospectively. For the current analysis, we selected nonrandomized patients who received facet joint or SI joint RF denervation and any form of exercise provided by a physiotherapist or exercise therapist. For patients in the observational study, this exercise program was not standardized in time or duration.

Outcomes

The three outcome measures were pain intensity (11-point Numerical Rating Scale [NRS]) [21], functional status (Oswestry Disability Index [ODI] 0-100) [[20], [21], [22]], and treatment success (global perceived effect [GPE], 7-point Likert scale) [23]. Treatment success measured by the GPE scale was defined as patients reporting to be "much improved" or "completely recovered."

Minimal clinically important change scores for patients with chronic LBP were estimated at 30% on the NRS for pain and 8 to 12 points on the ODI for functional status to be clinically relevant [24,25].

Age, gender, BMI, education, smoking habits, marital status, complaint history, and patient expectations were assessed on baseline. Patient expectations were assessed by the Credibility/Expectancy Questionnaire [26]. Health care utilization in primary and secondary care and the use of prescribed and over-the-counter medication were assessed by self-completed cost questionnaires [27].

All patients in the RCTs as well as observational study received the same questionnaires and were followed up for 12 months. For this study, we used the results up to the first 3 months of follow up, as we expected a randomization effect mostly during this 3-month intervention period. All questionnaires were web-based and sent at baseline and 3 months after start of treatment. Pain intensity and GPE were assessed at three and 6 weeks after start of treatment as well.

Statistical methods

Baseline characteristics of patients in the RCT intervention group and in the observational study were compared using descriptive statistics. We compared the randomized study groups pairwise to each of the two observational study groups, separately for LBP originating from the facet joints and SI joints. Baseline characteristics were compared between completers and noncompleters to identify possible selective dropout.

The analysis of mean changes for pain intensity included all available data for each of the predetermined follow-up assessments at 3 weeks, 6 weeks, and 3 months. To analyze differences between the groups in pain intensity, we used a maximum likelihood estimation from longitudinal linear mixed-effects models under "missing at random" assumptions, and included a fixed term for pain clinic if necessary, based on

the likelihood ratio test [28]. This fixed term for pain clinic was added to the model only for analyzing patients with facet joint pain.

For treatment success (dichotomous outcome), we used a generalized linear mixed model (logit link) with the same multilevel structure under "missing at random" assumptions and also included a fixed term for pain clinic if necessary, based on the likelihood ratio test [28]. This fixed term for pain clinic was added to the model for analyzing both patients with facet joint pain and SI joint pain.

Functional status was assessed at 3 months only. Analyzing differences in functional status was performed using a linear regression model.

For all analyses, we calculated regression coefficients or odds ratios with 95% confidence intervals and performed an unadjusted analysis and an analysis adjusted for baseline outcomes. We included time*group interactions in all models, and time-specific associations are presented, regardless of the statistical significance of these interaction terms. We used MLwiN to analyze the data (V2.22).

We preplanned a complete-case analysis as sensitivity analysis, including complete cases from the RCT (Facet RCT *N* = 98; SI joint RCT *N* = 87) and complete cases from the 3- and 6-week measurements in the observational groups (facet joint observational study group 1, *N* = 268; facet joint observational study group 2. *N* = 295; SI joint observational study group 1, *N* = 187; SI joint observational study group 2, *N* = 235). In the observational study groups, we selected patients based on self-reported RF denervation and physiotherapy. Self-reported treatments were assessed at the end of the 3-month online questionnaire. In this questionnaire, the outcomes were measured before the treatments and because participants were instructed to answer each question to continue the questionnaire, there were no missing data in any of the outcomes for the observational groups at 3 months.

RESULTS

In total, 7,529 patients were included in one of the RCTs or the observational group in the MinT study between January 1, 2013 and December 17, 2015. Most patients were excluded from the RCTs because of psychological problems (i.e., depressive symptoms or anxiety) or because they were older than 70 years; for a complete overview of exclusions, see Appendix A.

In total, 1,148 patients fulfilled the eligibility criteria for the present study (see Figure 1). Of the patients with LBP originating from the facet joints, 125 patients were randomized to the intervention group, 257 patients participated in observational group 1 (fulfilling all of the RCT eligibility criteria), and 233 patients participated in observational group 2 (not fulfilling the RCT eligibility criteria). Of the patients

with LBP originating from the SI joints, 116 patients were randomized to the intervention group, 198 patients in observational group 1, and 219 in observational study group 2.

Patient characteristics

Baseline characteristics are shown in Table 2. The total study population had a mean age of 55.6 years (SD 13.4), the majority was women, had a low education level, married, and had on average 13 years (SD 12.5) of LBP symptoms. During the 3-month follow-up, most of the patients visited primary care more than 10 times, visited the outpatient clinic at least once, and were never hospitalized (Table 3). On average, 20% of the patients used weak opioids and 4–18% used strong opioids (Table 3).

Patients in observational study group 2 were somewhat older, had a lower education level, less likely to have a paid job, had slightly more functional limitations at baseline, and more often used strong opioids compared with randomized patients and patients in observational study group 1 (Table 2, Table 3, Table 4). This applied to patients with chronic LBP originating from the facet joints and the SI joints.

There were hardly any differences between completers and noncompleters in baseline characteristics, health care use, medication, and outcomes (Appendix B and Appendix C). The number of complete cases ranged from 77% to 89% between the groups (Figure 1).

Comparison between randomised patients and patients in the observational study

Facet joints

The mean pain intensity for patients with LBP originating from the facet joints in the first 3 weeks after RF denervation decreased by 1.97 points (on an 11-point scale) in the randomized study group, 2.07 points in observational group 1, and 1.76 in observational group 2. Pain intensity stabilized afterward (Table 4 and Figure 2).

Comparing observational group 2 with the RCT

No statistical significantly or clinically relevant differences in improvement in pain intensity between patients in observational study group 2 and randomized patients were found. Statistically significantly less improvement in functional status in patients was shown in observational study group 2 compared with randomized patients (mean difference [MD] 5.82; 95% CI: 2.54–9.11) on a 0–100 scale. However, this difference is not considered clinically relevant (i.e., < 8 to 12 points on the ODI). We found a smaller

treatment success in observational study group 2 compared with the RCT overall and at each time point (three and 6 weeks, and 3 months after intervention) (Table 4).

SI-joints

The mean pain intensity for patients with LBP originating from the SI joints decreased 2.21 points in randomized patients in the first 3 weeks after receiving RF denervation, 2.31 points in observational group 1, and 2.05 points in observational group 2. This stabilized afterward (Table 4 and Figure 2).\

Comparing observational group 1 with the RCT

We found no differences in course of LBP symptoms based on pain, functional status, and treatment success between patients in observational study group 1 and randomized patients.

Comparing observational group 2 with the RCT

Participants in observational study group 2 had statistically significant less pain reduction overall (MD, 0.40; 95% CI: 0.09–0.72) on a 0–10 scale, and at the 3-month assessment as well (MD, 0.58; 95% CI 0.05–1.11). However, these MDs are small and not considered clinically relevant (i.e., < 30% on the NRS for pain intensity). Participants in observational study group 2 also had less improvement in functional status (MD, 7.16; 95% CI: 3.84–10.47) on a 0–100 scale over the 3-month period and overall less treatment success (OR, 0.72; CI: 0.54–0.96) compared with randomized patients. These differences are not considered clinically relevant either.

Sensitivity analysis

The percentage of patients with LBP originating from the facet joints who had complete data on all measurement points was almost 10% higher in the observational study compared with the randomized study groups (78.4% in the RCT vs. 88.3% in observational study group 1 and 89.0% in observational study group 1, respectively). Among patients with chronic LBP originating from the SI joints, the percentage of complete cases was 75.0% in the randomized study group and 76.8% and 77.6% in both study groups, respectively (see Figure 1).

We found merely negligible differences between the main analysis and complete case analysis (Appendix D).

Models without adjustments for baseline outcome differences (Appendix E) showed similar results in pain intensity and showed larger MDs between the observational study groups and the randomized

study groups in ODI scores. One explanation for this could be regression to the mean because the observational study groups started out with a higher pain intensity score and more limitations in functional status.

DISCUSSION

Main results

This study compared the symptom course between randomized study groups and observational study groups for patients with chronic LBP originating from the facet joints or SI joints who were treated with RF denervation and an exercise program. Our results suggest that these patients experience similar levels of pain and functioning and improvement whether randomized to a treatment group or undergoing treatment in an observational study as long as all patients fulfill eligibility criteria. When this is not the case (i.e., patients did not fulfill all the inclusion criteria), small but seemingly clinically irrelevant differences were observed; patients from observational studies show higher pain intensity score, more limitations in functional status, and less treatment success compared with randomized patients. Descriptive statistics showed that patients who did not fulfill all RCT inclusion criteria were somewhat older, had a lower education level, were less likely to have a paid job, had slightly more functional limitations at baseline, and more often used strong opioids compared with randomized patients and patients who did meet all RCT inclusion criteria. These patients were more likely to have a slightly worse symptom course, which might imply that the results of the RCT are a minor overestimation of the results in real life (which we consider a real-life combination of patients in observational groups 1 and 2).

Previous studies estimated minimal change scores for patients with LBP of 30% on the NRS for pain, and 8 to 12 points on the ODI for functional status to be clinically relevant [24,25]. All study groups in our study showed average changes less than 30% on the NRS for pain intensity and less than 10 points in functional status over time on the ODI. The differences between the randomized and observational study groups in the present study cannot be considered clinically relevant.

All in all, this study suggests that (1) the RCT results of the MinT trials reflect clinical practice in a similar population and (2) participants in the RCT show slightly better results compared with the observational group in clinical practice that does not meet all eligibility criteria. This study adds to the current literature that results of pragmatic RCTs (more specifically, the clinical course of patients in the intervention arm) in a secondary care setting for patients who received RF denervation at the pain clinic are comparable to the clinical course after this treatment for similar patients in daily clinical practice.

Strengths and limitations

A strength of the present study is the nationwide study design, which resulted in a large sample of patients who were recruited in routine clinical care. Second, patients who were randomly allocated to be treated with RF denervation and an exercise program could be compared with patients who received a similar treatment but had a treatment choice in the observational study groups. This increased the applicability of the results in clinical practice, which is probably the average of the outcomes in observational groups 1 and 2. In other words, it shows that the symptom course is probably not affected by randomization. This supports the generalizability of the RCT results.

A limitation of this study was the inability to select a proper control group that was more or less comparable to the control group in the RCT, using the observational data. The variety of physiotherapy treatments would have made it impossible to select a group of patients with a comparable treatment as the standardized exercise program in the RCT control group. Therefore, it was not possible to perform a nonrandomized comparison of RF denervation in addition to an exercise program vs. an exercise program alone. Various studies and one meta-analysis performed a nonrandomized treatment comparison in the field of LBP [[29], [30], [31], [32]]. These studies showed similar results in clinical course of LBP symptoms compared with the present study. However, these previous studies analyzed a variety of mostly conservative treatments, and clinical heterogeneity could potentially have influenced the results.

Second, a reference standard for diagnosing facet joint or SI joint pain is not available [14]. In this pragmatic study, diagnostic tests that are commonly applied in clinical practice were used. Controversy concerning the ideal threshold value of pain reduction in the diagnostic blocks exists. A 50% cutoff has most frequently been used in previous studies [14] and in clinical practice. Performing two or more independent diagnostic blocks decreases the false-positive rate but increases the number of false-negative blocks [33]. Furthermore, a clinical trial showed that multiple blocks are not cost-effective [31].

Third, other patients may present themselves at the pain clinic in daily practice compared with the patients we included in the present study. However, the inclusion criteria of the RCTs are in line with treatment criteria in daily practice, and more importantly, the vast majority of patients in the Netherlands who were treated with RF denervation during the inclusion period participated in the MinT study. This was the result of a governmental regulation that only patients who participated in the MinT study were reimbursed for the treatment costs by their health insurance company. For that reason, we expect that patients in this study reflect patients in clinical practice more closely.

Fourth, the relatively large attrition is a potential limitation. However, most of the attrition is explained by the selection of patients with diagnosed SI and facet joint pain who were included in the observational study after the RCT inclusion period was closed. However, there was a relatively large number of dropouts at 12 months. Although we did not define differences between the complete-case analysis and the intention-to-treat analysis using all data, it is possible that completers are different from noncompleters, which could have biased the results of the complete-case analyses.

Reflections

Systematic reviews in health care often ignore the evidence generated in observational studies, as the general assumption is that observational studies overestimate the effects of treatments tested in RCTs [2,34]. Moreover, observational findings are more likely to be biased and are based on studies that lack a comparable control group. The difference in results between RCTs and observational studies are usually attributed to differences in methodological quality. A recent meta-analysis provides evidence to the contrary [35]. This meta-analysis examined factors that explain heterogeneity in clinical outcomes in the field of LBP, of which study design is one of the factors. The authors concluded that other effect modifiers were frequently more powerful explanatory variables than study design. These factors included pain duration, involvement of workers' compensation, presence of spondylolisthesis, levels fused, and previous surgery [35]. The results of the meta-analysis are in line with our results and suggest that differences between RCTs and observational study results in the field of LBP are primarily attributable to clinical factors and not by the difference in study design. More data have become available that show similar clinical course results from observational studies and RCTs in fields outside [2,[36], [37], [38]], as well as inside the field of LBP [5,35]. The results of our study seem to support these findings.

We acknowledge differences in techniques between countries and settings. Needle size and placement, duration, and temperature of RF denervation could be some of these differences. We encourage researchers and clinicians in other countries or settings to evaluate whether these procedures reflect their daily practice. Second, researchers can consider matching (e.g., by propensity score matching or coarsened exact matching) on demographic characteristics of patients in observational studies and patients in RCTs to further investigate randomization bias.

Most LBP studies show small treatment effects and the clinical course in symptoms tend to improve in the first 6 weeks, reaching a plateau over the following 12 months [39,40]. Our study results are in line with these findings in previous literature. It seems more likely that differences between study results

can be attributed to study setting, population, intervention, or other discrepancies between studies and not to study design itself [41]. We encourage future research using data from routine clinical care (real-life data). This might be promising for evaluating effectiveness of clinical interventions in situations when performing an RCT is complex or unethical, but methodological challenges need to be addressed, such as, for example, confounding by indication.

Conclusion

Despite the belief that observational studies are assumed to overestimate the effects of interventions evaluated in RCTs, our study showed fairly similar outcomes in pain, functioning, and treatment success after RF denervation over a 3-month time in patients with chronic LBP originating from the facet joints and SI joints in a randomized study treatment population compared with similar patients from an observational study.

Contributors

MvT as principal investigator secured funding for the MinT-study and had overall responsibility for the management of the study. FH (as co-principal investigator) and BK collaborated in obtaining funding. FH, MvT, RO, GG, BK, AV, JJ, and EM were involved in drafting the original study protocol. FH and JK provided extra support for intervention development. JJ and EM (in collaboration with 16 cooperating hospitals) were responsible for data collection. EM performed the data cleaning and statistical analyses and wrote the initial draft of the manuscript. All authors contributed to, critically revised, and approved the final manuscript.

Acknowledgements

The authors thank the Netherlands Organisation for Health Research and Development (171202013), the Dutch Society for Anaesthesiology, and the Dutch health insurance companies for funding this study. We thank Sie Pei Lim for her assistance in the data cleaning. The authors would also like to thank all patients, physiotherapists, and everyone working on the MinT study at one of the participating pain clinics.

Registration nm

Dutch National Trial Register number: NTR3531

Protocol

Maas E, Juch J, Groeneweg J, Ostelo R, Koes B, Verhagen A, van Raamt M, Wille F, Huygen F, van Tulder M (2012) Cost-effectiveness of minimal interventional treatments for chronic mechanical low back pain: design of four randomised controlled trials with an economic evaluation. BMC Musculoskelet Disord 13 (1):260

References

1. Angus DC. Fusing Randomized Trials With Big Data: The Key to Self-learning Health Care Systems? JAMA 2015;314(8):767-768.
2. Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. New England Journal of Medicine 2000;342(25):1878-1886.
3. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. New England Journal of Medicine 2000;342(25):1887-1892.
4. Van Spall HG, Toren A, Kiss A, Fowler RA. Eligibility criteria of randomized controlled trials published in high-impact general medical journals: a systematic sampling review. Jama 2007;297(11):1233-1240.
5. Artus, M., van der Windt, D., Jordan, K. P. & Croft, P. R. The clinical course of low back pain: a meta-analysis comparing outcomes in randomised clinical trials (RCTs) and observational studies. BMC Musculoskelet Disord 15, 68 (2014).
6. Britton A, McKee M, Black N, McPherson K, Sanderson C, Bain C. Threats to applicability of randomised trials: exclusions and selective participation. Journal of Health Services Research & Policy 1999;4(2):112-121.
7. Sokka T, Pincus T. Most patients receiving routine care for rheumatoid arthritis in 2001 did not meet inclusion criteria for most recent clinical trials or american college of rheumatology criteria for remission. The Journal of Rheumatology 2003;30(6):113.
8. Concato J, Lawler EV, Lew RA, Gaziano JM, Aslan M, Huang GD. Observational methods in comparative effectiveness research. The American journal of medicine 2010;123(12):e16-e23.
9. Colditz GA. Overview of the epidemiology methods and applications: strengths and limitations of observational study designs. Critical reviews in food science and nutrition 2010;50(S1):10-12.
10. Furlan AD, Tomlinson G, Jadad AAR, Bombardier C. Examining heterogeneity in meta-analysis: comparing results of randomized trials and nonrandomized studies of interventions for low back pain. Spine 2008;33(3):339-348.
11. Jacobs WC, Kruyt MC, Verbout AJ, Oner FC. Spine surgery research: on and beyond current strategies. The Spine Journal 2012;12(8):706-713.
12. Maas, E. T. *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.* **13,** 260 (2012).
13. Cosman Jr ER, Gonzalez CD. Bipolar radiofrequency lesion geometry: implications for palisade

treatment of sacroiliac joint pain. Pain Practice 2011;11(1):3-22.
14. Kline M. Radiofrequency techniques in clinical practice. Interventional Pain Management 2000.
15. Cohen, S. P., Huang, J. H. Y. & Brummett, C. Facet joint pain--advances in patient selection and treatment. *Nat. Rev. Rheumatol.* **9,** 101–16 (2013).
16. Cohen, S. P. & Raja, S. N. Zygapophysial (Facet) Joint Pain. 591–614 (2007).
17. Juch, J. N. S. *et al.* Effect of Radiofrequency Denervation on Pain Intensity Among Patients With Chronic Low Back Pain. *Jama* **318,** 68 (2017).
18. Cohen, S. P. *et al.* Randomized placebo-controlled study evaluating lateral branch radiofrequency denervation for sacroiliac joint pain. *Anesthesiology* **109,** 279–88 (2008).
19. Schmidt PC, Pino CA, Vorenkamp KE. Sacroiliac joint radiofrequency ablation with a multilesion probe: A case series of 60 patients. Anesthesia & Analgesia 2014;119(2):460-462.
20. Downie, W. W. *et al.* Studies with pain rating scales. *Ann. Rheum. Dis.* **37,** 378–381 (1978).
21. Fairbank, J. C. T. Why are there different versions of the Oswestry Disability Index? *J. Neurosurg. Spine* **20,** 83–86 (2014).
22. Kamper, S. J. *et al.* 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,** 760–766.e1 (2010).
23. Devilly, G. J. & Borkovec, T. D. Psychometric properties of the credibility/expectancy questionnaire. *J. Behav. Ther. Exp. Psychiatry* **31,** 73–86 (2000).
24. Goossens, M. E., Rutten-van Mölken, M. P., Vlaeyen, J. W. & van der Linden, S. M. The cost diary: a method to measure direct and indirect costs in cost-effectiveness research. *J. Clin. Epidemiol.* **53,** 688–95 (2000).
25. Twisk J. Applied multilevel analysis: a practical guide for medical researchers: Cambridge Univ Press, 2006.
26. Hägg O, Fritzell P, Nordwall A. The clinical importance of changes in outcome scores after treatment for chronic low back pain. European Spine Journal 2003;12(1):12-20.
27. Ostelo RW, de Vet HC. Clinically important outcomes in low back pain. Best practice & research clinical rheumatology 2005;19(4):593-607.
28. Bonetti F, Curti S, Mattioli S, Mugnai R, Vanti C, Violante FS, Pillastrini P. Effectiveness of a'Global Postural Reeducation'program for persistent low back pain: a non-randomized controlled trial. BMC musculoskeletal disorders 2010;11(1):285.
29. Diamond TH, Bryant C, Browne L, Clark WA. Clinical outcomes after acute osteoporotic vertebral

fractures: a 2-year non-randomised trial comparing percutaneous vertebroplasty with conservative therapy. Medical Journal of Australia 2006;184(3):113.

30. Gerszten P, Welch W, McGrath P, Willis S. A prospective outcomes study of patients undergoing intradiscal electrothermy (IDET) for chronic low back pain. Pain Physician 2002;5(4):360-364.

31. Weinstein, J. N. *et al.* Surgical vs Nonoperative Treatment for Lumbar Disk Herniation. **296,** 2441–2450 (2006).

32. Kunz R, Oxman AD. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. Bmj 1998;317(7167):1185-1190.

33. Anglemyer A, Horvath HT, Bero L. Healthcare outcomes assessed with observational study designs compared with those assessed in randomized trials. Cochrane Database Syst Rev 2014;4.

34. Golder S, Loke YK, Bland M. Meta-analyses of adverse effects data derived from randomised controlled trials as compared to observational studies: methodological overview. PLoS Med 2011;8(5):e1001026.

35. Ioannidis JP, Haidich A-B, Pappa M, Pantazis N, Kokori SI, Tektonidou MG, Contopoulos-Ioannidis DG, Lau J. Comparison of evidence of treatment effects in randomized and nonrandomized studies. Jama 2001;286(7):821-830.

36. Artus M, van der Windt DA, Jordan KP, Hay EM. Low back pain symptoms show a similar pattern of improvement following a wide range of primary care treatments: a systematic review of randomized clinical trials. Rheumatology 2010:keq245.

37. Hestbaek L, Leboeuf-Yde C, Manniche C. Low back pain: what is the long-term course? A review of studies of general patient populations. European Spine Journal 2003;12(2):149-165.

38. Shrier, I. *et al.* Should meta-analyses of interventions include observational studies in addition to randomized controlled trials? A critical examination of underlying principles. *Am. J. Epidemiol.* **166,** 1203–1209 (2007).

Tables and figures:

- Figure 1. Flow chart
- Table 1. Descriptive baseline characteristics – CLBP originating from the facet joints and the SI-joints
- Table 2. Descriptive healthcare and medication use – LBP originating from the facet joints and the SI-joints
- Table 3. Outcomes of the observational study groups compared to the randomised study groups
- Figure 2. Course in low back pain symptoms at baseline until three months follow-up

Appendices:

- Appendix 1. Reasons for exclusion from the RCT
- Appendix 2. Complete case descriptive baseline characteristics – CLBP originating from the facet joints and the SI-joints
- Appendix 3. Complete case descriptive healthcare and medication use – CLBP originating from the facet joints and the SI-joints
- Appendix 4. Complete case analysis: Treatment effects of observational study groups compared to the RCT

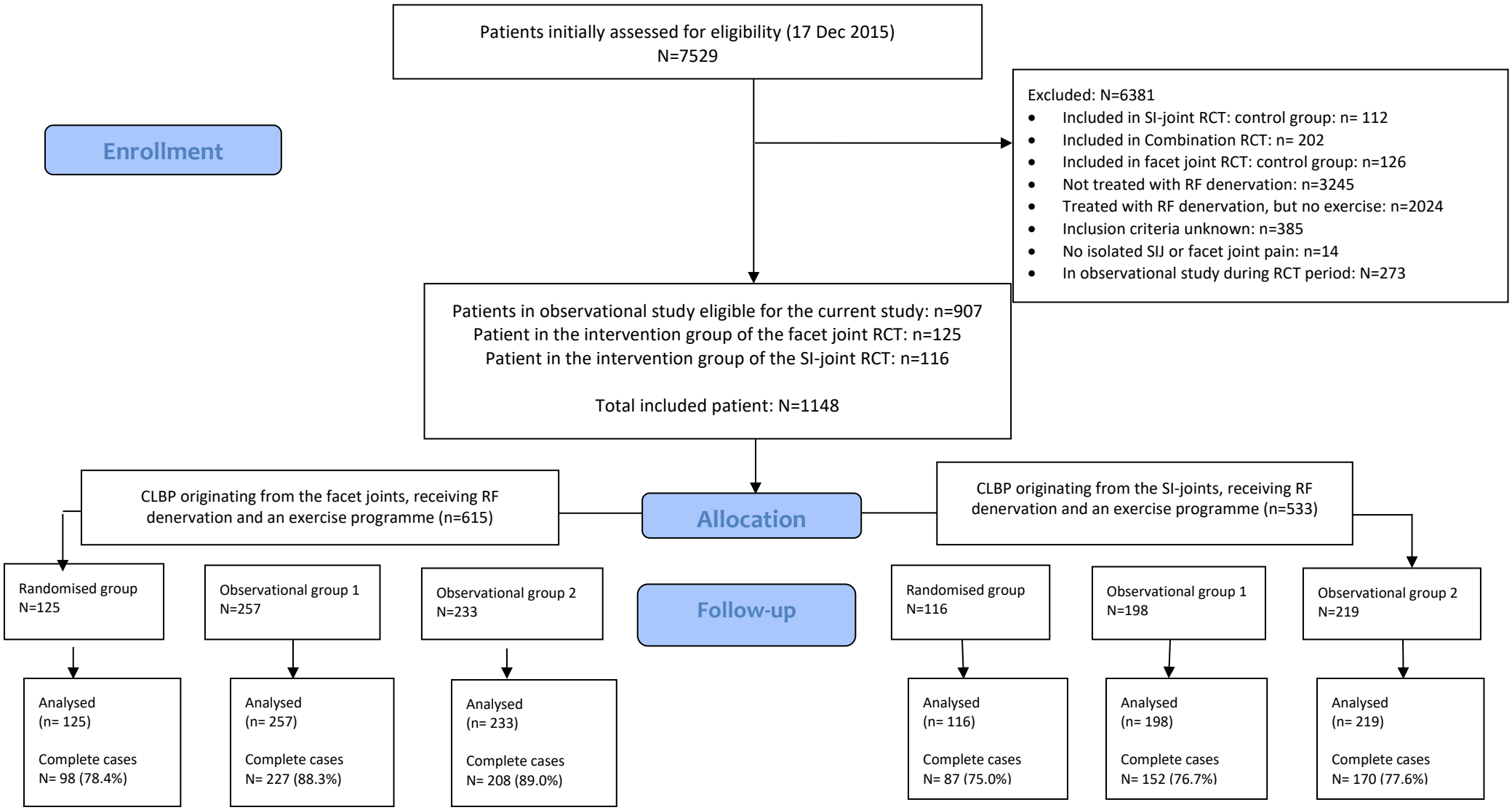


FIGURE 1. FLOW CHART

TABLE 1. DESCRIPTIVE BASELINE CHARACTERISTICS – CLBP ORIGINATING FROM THE FACET JOINTS AND THE SI-JOINTS

Characteristics	Patients with CLBP originating from the facet joints[A]			Patients with CLBP originating from the SI-joints[A]		
	Randomised group N=125	Observational group 1 N=257	Observational group 2 N=233	Randomised group N=116	Observational group 1 N=198	Observational group 2 N=219
Age, mean (SD), y	52.9 (11.5)	53.3 (11.04)	62.7 (13.91)	51.6 (10.9)	51.4 (11.0)	57.4 (15.6)
Women, No. (%)	65 (55.6%)	163 (63.4%)	144 (61.8%)	87 (75.0%)	150 (79.8%)	165 (78.2%)
BMI, mean (SD)	26.7 (5.2)	27.1 (3.8)	29.0 (5.8)	26.7 (4.2)	26.5 (3.8)	28.9 (6.3)
Smoker, No. (%)	34 (29.1%)	67 (26.1%)	51 (21.9%)	29 (26.6%)	42 (22.3%)	50 (23.7%)
Education level, No, (%)[B]						
• **Low**	57 (48.7%)	116 (45.1%)	136 (58.4%)	59 (54.1%)	80 (42.6%)	119 (56.9%)
• **Moderate**	35 (29.9%)	80 (31.1%)	50 (21.5%)	32 (29.4%)	74 (39.4%)	62 (29.7%)
• **High**	21 (17.9%)	61 (23.7%)	47 (20.2%)	18 (16.5%)	34 (18.1%)	28 (13.4%)
History of back pain complaints, median (IQR), months						
• **Time since first experience with low back pain**	146 (50-267)	122 (37-244)	137 (50-266)	97 (37 -228)	122(37 – 241)	164 (37 – 244)
• **Time since first current episode with low back pain**	31 (12-103)	30 (12-67)	48 (13-122)	30 (12 – 76)	30 (12 – 85)	30 (12 – 73)
Married, or living with a partner, No. (%)	93 (74.4%)	215 (83.7%)	166 (71.2%)	85 (78.0%)	143 (76.1%)	143 (68.1%)
CEQ score, mean (SD)[C]						
• **Credibility (0-27)**	21.4 (3.9)	21.4 (4.2)	21.4 (3.9)	21.4 (4.5)	22.6 (3.2)	21.4 (4.0)
• **Expectancy (0-27)**	18.9 (4.6)	18.7 (4.7)	18.1 (4.6)	18.8 (4.9)	19.5 (4.3)	18.2 (4.8)
Having a paid job (%)	64 (51.2%)	144 (56.3%)	47 (20.2%)	66 (61.7%)	107 (54.0%)	70 (32.6%)

Abbreviations: SD, Standard Deviation; No, number; BMI, Body Mass Index (calculated as weight in kilograms divided by height in meters squared); IQR: Inter Quartile Range; SD: Standard Deviation; CEQ, credibility expectancy questionnaire

[A] Results are presented of the patients who had complete baseline data

[B] Education levels: Low indicated preschool, primary school or lower secondary school; moderate indicates higher secondary school or undergraduate; high indicates tertiary, university, or postgraduate.

[C] A higher score indicates more credibility in the effectiveness of treatment or higher expectations about the treatment (score range, 0-27)

Table 2. Descriptive healthcare and medication use – LBP originating from the facet joints and the SI-joints

Characteristics	Patients with CLBP originating from the facet joints			Patients with CLBP originating from the SI-joints		
	Randomised group N=125	Observational group 1 N=257	Observational group 2 N=233	Randomised group N=116	Observational group 1 N=198	Observational group 2 N=219
Primary care visits						
• **0 (%)**	12 (10.1%)	0 (0.0%)	0 (0.0%)	0 (0%)	0 (0.0%)	0 (0.0%)
• **<10 (%)**	27 (22.7%)	108 (42.0%)	94 (40.3%)	32 (29.4%)	95 (48.0%)	98 (44.7%)
• **≥10 (%)**	80 (67.2%)	149 (58.0%)	139 (59.7%)	77 (70.6%)	103 (52.0%)	121 (55.3%)
Outpatient clinic						
• **0 (%)**	49 (41.2%)	79 (30.7%)	71 (30.5%)	43 (39.4%)	73 (36.9%)	70 (32.0%)
• **≥1 (%)**	70 (58.8%)	178 (69.3%)	162 (62.5%)	66 (56.9%)	125 (63.1%)	149 (68.0%)
One day treatment						
• **0 (%)**	77 (64.7%)	141 (54.9%)	145 (62.2%)	62 (56.9%)	117 (59.1%)	132 (60.3%)
• **≥1 (%)**	42 (35.3%)	116 (45.1%)	88 (37.8%)	47 (43.1%)	81 (40.9%)	87 (39.7%)
Hospitalisation						
• **0 (%)**	119 (100.0%)	252 (98.1%)	230 (98.7%)	108 (99.1%)	193 (97.5%)	218 (99.5%)
• **≥1 (%)**	0 (0.0%)	5 (1.9%)	3 (1.3%)	1 (0.9%)	5 (2.5%)	1 (0.5%)
Medication use						
• **None/non back pain related**	46 (39.7%)	77 (30.0%)	59 (25.3%)	37 (34.3%)	54 (27.3%)	60 (27.4%)
• **Non-opioids (%) (aspirin/paracetamol/NSAIDs)**	42 (36.2%)	104 (40.5%)	77 (33.0%)	53 (38.9%)	81 (40.9%)	78 (35.6%)
• **Weak opioids (%) (with or without non-opioids)**	23 (19.8%)	46 (17.9%)	54 (23.2%)	24 (22.2%)	41 (20.7%)	48 (21.9%)
• **Strong opioids (%) (with or without non-opioids)**	5 (4.3%)	28 (10.9%)	41 (17.6%)	5 (4.6%)	21 (10.6%)	29 (13.2%)

Abbreviations: NSAIDS: Nonsteroidal anti-inflammatory drugs

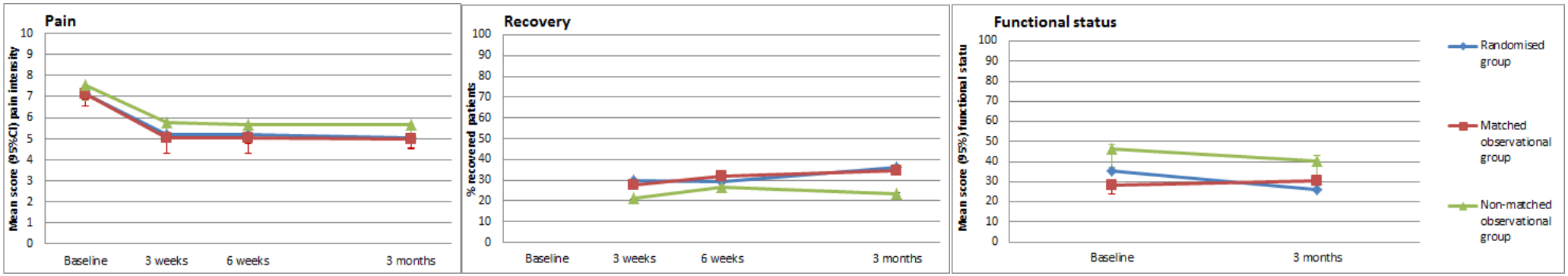
Table 3. Outcomes of the observational study groups compared to the randomised study groups

		Patients with CLBP originating from the facet joints					Patients with CLBP originating from the SI-joints				
		Randomised group N=125	Observational group 1 N=257	Mean difference (95%CI)	Observational group 2 N=233	Mean difference (95%CI)	Randomised group N=116	Observational group 1 N=198	Mean difference (95%CI)	Observational group 2 N=219	Mean difference (95%CI)
NRS Pain (SD)	Overall difference			-0.17 (-0.57-0.22)		0.35 (-0.05-0.75)			-0.03 (-0.35-0.30)		0.40 (0.09-0.72)
	Baseline	7.14 (1.38)	7.11 (1.395)		7.55 (1.38)		7.17 (1.65)	7.56 (1.33)		7.46 (1.44)	
	3 weeks	5.17 (2.27)	5.04 (2.156)	-0.28 (-0.76-0.21)	5.79 (2.09)	0.29 (-0.20-0.79)	4.96 (2.19)	5.25 (2.20)	0.13 (-0.45-0.72)	5.41 (2.34)	0.38 (-0.19-0.95)
	6 weeks	5.19 (2.31)	5.02 (2.25)	-0.22 (-0.69-0.25)	5.66 (1.99)	0.26 (-0.22-0.74)	5.22 (2.16)	4.89 (2.23)	-0.45 (-0.99-0.09)	5.60 (2.26)	0.25 (-0.28-0.78)
	3 months	5.01 (2.29)	5.00 (2.350)	-0.04 (-0.50-0.43)	5.68 (2.13)	0.40 (-0.11-0.92)	4.77 (2.46)	5.17 (2.38)	0.27 (-0.27-0.80)	5.45 (2.27)	0.58 (0.05-1.11)*
ODI Functioning (SD)	Baseline	35.08 (14.66)	28.24 (14.05)		46.07 (14.89)		38.07 (14.07)	41.01 (13.39)		46.22 (13.73)	
	3 months	26.03 (16.58)	30.53 (16.57)	1.47 (-1.67-4.60)	40.21 (17.55)	5.82 (2.54-9.11)	27.72 (17.05)	31.33 (15.01)	1.98 (-1.33-5.29)	39.94 (16.95)	7.16 (3.84-10.47)
				OR (95%CI)		**OR (95%CI)**			**OR (95%CI)**		**OR (95%CI)**
Treatment success (%)	Overall difference			0.99 (0.75-1.3)		0.67 (0.50-0.90)			0.93 (0.70-1.25)		0.72 (0.54-0.96)
	3 weeks	32 (29.6%)	63 (27.5%)	0.90 (0.54-1.49)	45 (21.1%)	0.64 (0.38-1.08)	28 (29.8%)	47 (29.7%)	0.98 (0.56-1.72)	47 (26.4%)	0.83 (0.48-1.45)
	6 weeks	35 (29.4%)	81 (32.1%)	1.14 (0.71-1.83)	60 (26.7%)	0.87 (0.53-1.43)	40 (37.0%)	63 (33.2%)	0.84 (0.51-1.38)	61 (28.5%)	0.68 (0.42-1.11)
	3 months	43 (36.1%)	89 (34.6%)	0.94 (0.59-1.48)	55 (23.6%)	0.55 (0.34-0.88)	43 (39.1%)	77 (38.9%)	0.98 (0.61-1.57)	67 (30.6%)	0.68 (0.42-1.09)

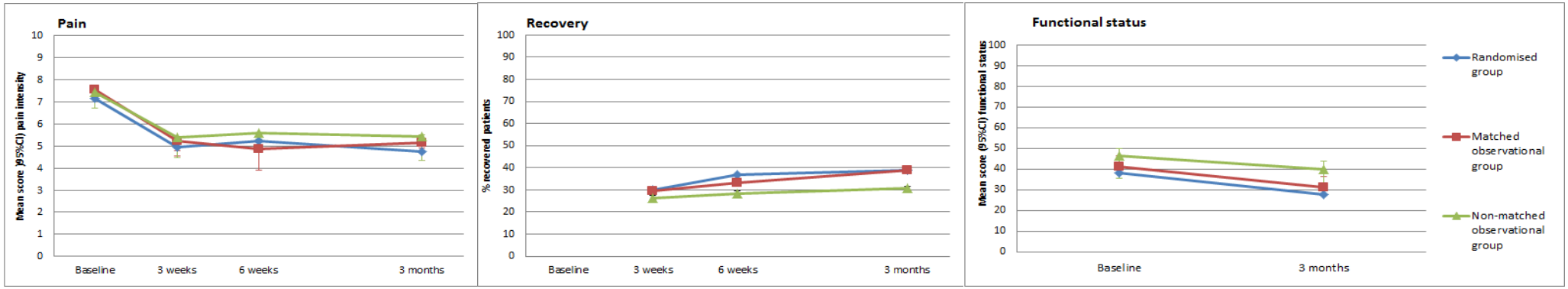
Values presented are means with corresponding standard deviations (SD), percentages of recovered patients, and model estimates of linear mixed-effects models with a random intercept, adjusted for pain intensity and functional status at baseline. Regression coefficients can be interpreted as mean differences between both observational groups compared to the RCT at a certain follow-up moment compared to baseline. Abbreviations: OR. Odds Ratio; NRS. Numeric Rating Scale (0-10); GPE. Global Perceived Effect; ODI. Oswestry Disability Index (0-100).
Higher score indicates more severe symptoms.

Figure 2. Course in low back pain symptoms at baseline until three months follow-up

1. Patients with chronic LBP originating from the facet joints



2. Patients with chronic LBP originating from the SI-joints



Values presented are unadjusted mean outcome score at baseline and each follow-up moment. Abbreviation: NRS. Numeric Rating Scale (0-10); GPE. Global Perceived Effect; ODI. Oswestry Disability Index (0-100). Higher score indicates more severe symptoms.* P<0.05 compare to RCT intervention group as reference

APPENDIX 1 REASONS FOR EXCLUSION FROM THE RCT

Exclusion criteria	Patients with CLBP originating from the facet joints RCT N=233	Patients with CLBP originating from the SI-joints RCT N=219
Psychological problems	88 (37.8%)	108 (49.3%)
Age >70	58 (24.9%)	32 (14.6%)
Psychological problems & age>70	24 (10.3%)	17 (7.8%)
BMI>35	21 (9.0%)	22 (10.0%)
Negative diagnostic block	16 (6.9%)	20 (9.1%)
Psychological problems & BMI>35	11 (4.7%)	8 (3.7%)
Psychological problems & negative diagnostic block	5 (2.1%)	7 (3.2%)
Negative diagnostic block & age>70	3 (1.3%)	0 (0.0%)
Age>70 & BMI>35	2 (0.9%)	1 (0.5%)
Psychological problems & age>70 & BMI>35	1 (0.4%)	2 (0.9%)
Negative diagnostic block & BMI>35	2 (0.9%)	0 (0.0%)
Psychological problems & BMI>35 & negative diagnostic block	2 (0.9%)	1 (0.5%)
Psychological problems & age>70 & negative diagnostic block	0 (0.0%)	1 (0.5%)

Abbreviations: BMI: Body Mass Index

Appendix 2. Complete case descriptive baseline characteristics – CLBP originating from the facet joints and the SI-joints

Characteristics	Patients with CLBP originating from the facet joints[A]			Patients with CLBP originating from the SI-jointsA		
	Randomised group N=98	Observational group 1 N=227	Observational group 2 N=208	Randomised group N=87	Observational group 1 N=152	Observational group 2 N=170
Age in years (SD)	53.6 (11.3)	53.6 (11.1)	62.7 (13.8)	51.9 (10.9)	50.9 (11.1)	57.4 (15.3)
Female	54 (55.1%)	143 (63.0%)	127 (61.1%)	70 (80.5%)	124 (81.6%)	129 (75.9%)
BMI (SD)	26.6 (5.2)	27.2 (3.8)	29.0 (5.7)	26.5 (4.2)	26.3 (3.8)	29.0 (6.5)
Smoker	29 (29.6%)	59 (26.0%)	47 (22.6%)	25 (28.7%)	34 (22.4%)	43 (25.3%)
Education [B]						
• **Low**	48 (49.0%)	106 (46.7%)	124 (59.6%)	49 (56.3%)	66 (42.8%)	96 (56.5%)
• **Moderate**	29 (29.6%)	69 (30.4%)	44 (21.2%)	25 (28.7%)	60 (39.5%)	49 (28.8%)
• **High**	21 (21.4%)	52 (22.9%)	40 (19.2%)	13 (14.9%)	26 (17.1%)	24 (14.1%)
History of back pain complaints						
• **Months first LBP experience (median (IQR))**	146 (52-271)	122 (37-243)	194 (52-291)	120 (37-231)	115 (37-115)	97 (37-282)
• **Months with current LBP episode (median (IQR))**	30 (12-97)	28 (12-67)	49 (14-123)	30 (10-79)	28 (12-85)	33 (12-73)
Married	77 (78.6%)	189 (83.3%)	146 (70.2%)	70 (80.5%)	117 (77.0%)	119 (69.4%)
Expectations [C]						
• **Credibility (0-27)**	21.5 (2.59)	21.4 (4.2)	21.4 (4.0)	21.1 (4.5)	22.9 (3.2)	21.4 (3.9)
• **Expectancy (0-27)**	19.2 (4.4)	18.6 (4.8)	18.2 (4.6)	18.5 (4.9)	19.6 (4.4)	18.3 (4.7)
Having a paid job	52 (53.6%)	130 (57.3%)	41 (19.7%)	51 (60.0%)	87 (57.2%)	58 (34.1%)

Abbreviations: SD, Standard Deviation; No, number; BMI, Body Mass Index (calculated as weight in kilograms divided by height in meters squared); IQR: Inter Quartile Range; SD: Standard Deviation; CEQ, credibility expectancy questionnaire

[A] Results are presented of the patients who had complete data

[B] Education levels: Low indicated preschool, primary school or lower secondary school; moderate indicates higher secondary school or undergraduate; high indicates tertiary, university, or postgraduate.

[C] A higher score indicates more credibility in the effectiveness of treatment or higher expectations about the treatment (score range, 0-27)

APPENDIX 3. COMPLETE CASE DESCRIPTIVE HEALTHCARE AND MEDICATION USE – CLBP ORIGINATING FROM THE FACET JOINTS AND THE SI-JOINTS

	Patients with CLBP originating from the facet joints			Patients with CLBP originating from the SI-joints		
Characteristics	Randomised group N=98	Observational group 1 N=227	Observational group 2 N=208	Randomised group N=87	Observational group 1 N=152	Observational group 2 N=170
Primary care visits						
• **0**	9 (9.2%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
• **<10**	20 (20.4%)	94 (41.1%)	83(39.9%)	25 (28.7%)	71 (46.7%)	75 (44.1%)
• **≥10**	69 (70.4%)	133 (58.6%)	125 (60.1%)	62 (71.3%)	81 (53.3%)	95 (55.9%)
Outpatient clinic						
• **0**	40 (40.8%)	72 (31.7%)	62 (29.8%)	35 (40.2%)	54 (35.5%)	55 (32.4%)
• **≥1**	58 (59.2%)	155 (68.3%)	146 (70.2%)	52 (59.8%)	98 (64.5%)	115 (67.6%)
One day treatment						
• **0**	65 (66.3%)	120 (52.9%)	128 (61.5%)	47 (54.0%)	88 (57.9%)	104 (61.2%)
• **≥1**	33 (33.7%)	107 (47.1%)	80 (38.5%)	40 (46.0%)	64 (42.1%)	66 (38.8%)
Hospitalisation						
• **0**	98 (100%)	223 (98.2%)	205 (98.6%)	86 (98.9%)	149 (98.0%)	169 (99.4%)
• **≥1**	0 (0.0%)	4 (1.8%)	3 (1.4%)	1 (1.1%)	3 (2.0%)	1 (0.6%)
Medication use						
• **None/non back pain related**	36 (36.7%)	68 (30.0%)	53 (25.5%)	32 (37.2%)	45 (29.6%)	46 (27.1%)
• **Non-opioids (aspirin/paracetamol/NSAID)**	38 (39.2%)	91 (40.1%)	66 (31.7%)	31 (36.0%0	60 (39.5%)	67 (39.4%)
• **Weak opioids (with or without non-opioids)**	19 (19.6%)	40 (17.6%)	49 (23.6%)	19 (21.8%)	28 (18.4%)	36 (21.2%)
• **Strong opioids (with or without non-opioids)**	4 (4.1%)	26 (11.5%)	38 (18.3%)	4 (4.7%)	18 (11.8%)	19 (11.2%)

APPENDIX 4. COMPLETE CASE ANALYSIS: TREATMENT EFFECTS OF OBSERVATIONAL STUDY GROUPS COMPARED TO THE RCT

		Patients with CLBP originating from the facet joints					Patients with CLBP originating from the SI-joints				
		Randomised group N=98	**Observational group 1 N=227**	**B (95%CI)**	**Observational group 2 N=208**	**B (95%CI)**	**Randomised group N=87**	**Observational group 1 N=152**	**B (95%CI)**	**Observational group 2 N=170**	**B (95%CI)**
NRS Pain (SD)	Overall effect			-0.03 (-0.45-0.38)		0.49 (0.07-0.92)*			0.01 (-0.49-0.51)		0.34 (-0.14-0.83)
	Baseline	7.10 (1.41)	7.12 (1.35)		7.52 (1.41)		7.03 (1.76)	7.67 (1.17)		7.41 (1.43)	
	3 weeks	5.20 (2.22)	5.05 (2.16)	-0.16 (-0.66-0.34)	5.76 (2.09)	0.40 (-0.11-0.91)	5.01 (2.21)	5.22 (2.23)	0.08 (-0.52-0.69)	5.43 (2.35)	0.35 (-0.24-0.93)
	6 weeks	5.12 (2.33)	5.04 (2.56)	-0.09 (-0.60-0.41)	5.66 (1.96)	0.38 (-0.13-0.91)	5.21 (2.17)	4.99 (2.19)	-0.34 (-0.94-0.27)	5.51 (2.33)	0.23 (-0.36-0.82)
	3 months	4.87 (2.25)	5.03 (2.36)	0.16 (-0.35-0.66)	5.72 (2.13)	0.70 (0.19-1.21)	4.74 (2.54)	5.14 (2.39)	0.29 (-0.32-0.89)	5.26 (2.29)	0.45 (-0.14-1.04)
ODI Functioning (SD)	Baseline	35.33 (14.49)	38.28 (13.88)		46.18 (15.02)		38.51 (14.05)	40.71 (13.33)		45.32 (13.84)	
	3 months	26.22 (17.12)	30.78 (16.71)	2.39 (-0.88-5.56)	40.25 (17.68)	6.26 (2.84-9.68)	27.10 (17.51)	30.15 (14.44)	1.61 (-2.03-5.25)	38.04 (16.98)	6.72 (3.10-10.35)
				OR (95%CI)		**OR (95%CI)**			**OR (95%CI)**		**OR (95%CI)**
Treatment success	Overall effect			0.95 (0.63-1.44)		0.63 (0.41-0.97)*			-0.10 (-0.55-0.36)		-0.28 (-0.73-0.17)
	3 weeks	30 (30.6%)	62 (27.3%)	0.85 (0.47-1.55)	45 (21.6%)	0.63 (0.34-1.16)	26 (29.9%)	47 (30.9%)	0.05 (-0.61-0.71)	45 (26.5%)	-0.17 (-0.83-0.49)
	6 weeks	29 (29.6%)	74 (32.6%)	1.15 (0.63-2.09)	54 (26.0%)	0.83 (0.45-1.54)	34 (39.1%)	49 (32.2%)	-0.30 (-0.94-0.34)	52 (30.6%)	-0.38 (-1.01-0.26)
	3 months	37 (37.8%)	79 (34.8%)	0.88 (0.49-1.57)	47 (22.6%)	0.48 (0.26-0.88)	35 (40.2%)	60 (39.5%)	-0.03 (-0.66-0.60)	57 (33.5%)	0.29 (-0.34-0.91)

Values presented are model estimates of linear mixed-effects models with a random intercept, and adjusted for outcomes at baseline. Regression coefficients can be interpreted as mean differences between interventions at a certain follow-up moment compared to baseline. Abbreviation: OR. Odds Ratio; NRS. Numeric Rating Scale (0-10); GPE. Global Perceived Effect; ODI. Oswestry Disability Index (0-100). Higher score indicates more severe symptoms.