8

Do randomised controlled trials reflect clinical practice?

Comparing the course of low back pain symptoms in randomised study groups to observational study groups

Esther T Maas, Johan NS Juch, Raymond WJG Ostelo, J George Groeneweg,
Jan-Willem Kallewaard, Michiel R de Boer, Bart Koes, Arianne P Verhagen,
Frank JPM Huygen, Maurits W van Tulder
In preparation for submission

ABSTRACT

Introduction There is an on-going debate whether or not randomised trial results reflect results in observational studies, which are presumed to reflect daily practice more closely. Two randomised controlled trials (RCTs) assessed the effectiveness of radiofrequency (RF) denervation added to a standardised exercise programme for patients with chronic low back pain (CLBP) originating from the facet joints or sacroiliac (SI-) joints. These RCTs were conducted alongside an observational prospective cohort study including comparable patients. The aim of the current study is to assess the generalizability of the results from these RCTs by comparing the three-month course in LBP symptoms between the randomised study groups and the observational study groups.

Methods Eligible patients had CLBP originating from the facet joints or SI-joints, had not responded to conservative care, and were included in the RCTs or the observational study. The matched observational study groups consisted of patients with the same inclusion criteria as patients in the RCTs. The non-matched observational study groups consisted of patients who were excluded from the RCTs because they did not meet one or more of the inclusion criteria. All patients were treated with RF denervation added to an exercise programme. Outcomes were the course in pain intensity (numeric rating scale 0-10), global perceived recovery (7-point Likert scale; treatment success was achieved if a patient reported being “much improved” or “complete recovery”), and functional status (Oswestry Disability Index 0-100) during three months after the RF denervation. Longitudinal mixed-model analyses for the differences in outcomes between the randomised and observational study groups were performed.

Results In total, 7529 participants were included in one of the RCTs or the observational study, and 1421 participants fulfilled the selection criteria for the current study. From the patients with LBP originating from the facet joints, 125 patients were randomised in the intervention group, 307 patients were included in the matched and 297 patients were included in the non-matched observational study group. From the patients with LBP originating from the SI-joints, 116 patients were included in the randomised study group, 243 patients in the matched and 297 in the non-matched observational study group. No statistical significant differences in the course of pain intensity, functional status, or recovery were found between participants with facet joint and SI-joint pain in the matched observational study group compared to the randomised study group. Participants in the non-matched observational study group showed statistically

significant less improvement in functional status compared to the participants in the randomised study group for patients with facet joint pain (mean difference (MD) 4.34; 95% confidence interval (CI): 1.42-7.25) and SI-joint pain (MD 6.41; 95%CI: 3.30-9.51), on a 0-100 scale. Participants with SI-joint pain in the non-matched observational study group showed statistically significant less decrease in pain intensity compared to the randomised group as well (MD 0.46; 95%CI: 0.04-0.88), on a 0-10 scale.

Conclusions Overall, we found no clinically relevant differences in course of LBP symptoms between randomised study groups and observational study groups for patients with CLBP originating from the facet joints or SI-joints who were treated with RF denervation and an exercise programme. Results from the RCTs seem to be generalizable to a similar patient population in clinical practice, and only small differences were shown when comparing randomised study groups to observational study groups with a slightly different population.

INTRODUCTION

There is an on-going debate whether randomised trial results might not reflect the results of treatment in observational studies, in turn are presumed to reflect daily practice more closely. In systematic reviews, the evidence generated in observational studies is often ignored because of the assumption that their findings might be biased.[1] Although systematic reviews of randomised controlled trials (RCTs) generally provide the strongest evidence for the (comparative) effectiveness of treatments, it has also been suggested that participation in a randomised trial might influence the course of symptoms.[2] The willingness of patients to be randomly allocated to a treatment might make these individuals different from the average patient.[2] This raises the question whether participants in RCTs are less representative to the average patients compared with participants in observational studies in which participants are not randomised. It is therefore important to explore to what extend the outcomes and clinical course in symptoms in patients included in the two study designs are comparable.

Two recent RCTs evaluated the effectiveness of radiofrequency (RF) denervation added to a standardised exercise programme for patients with CLBP originating from the facet joints and sacroiliac (SI)-joints. 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.[3,4] RF denervation is a commonly used treatment in patients with LBP originating from the facet joints or SI-joints.[5,6] This treatment is provided in pain clinics and (usually) part of a multidisciplinary treatment programme. Both RCTs were part of a larger collective initiative, to evaluate RF denervation for patients with CLBP: the MinT (Minimal Invasive Treatment) study.7 The MinT study consisted of an observational cohort and three RCTS (for patients with facet joint pain, SI-joint pain, and a combination of facet joint, SI-joint, and/or disc pain). The MinT study provides an excellent opportunity to compare results of RCT data with observational data, because the vast majority of patients in the Netherlands who were treated with RF denervation during the inclusion period participated in the MinT study.

The aim of the current study is to assess the generalizability of the results from these RCTs by comparing the course in LBP symptoms over a three-month time period between randomised study groups and observational study groups.

METHODS

Study design and setting

The MinT study was a nationwide, multicentre study, conducted in 16 pain clinics and 102 physiotherapy practices in the Netherlands. In the published study protocol, four RCTs were described.[7] The RCT in which participants with isolated discogenic problems were assessed was prematurely terminated, as no participants were diagnosed after an inclusion period of five months. The Medical Ethics Committee of the Erasmus University Medical Centre in Rotterdam granted ethical approval (registration number MEC-2012-079). All included participants gave written informed consent.

Participants

Eligible participants in the observational study had CLBP, showed no improvement of symptoms after conservative treatment, were referred to a pain clinic and were able to complete Dutch questionnaires.

Extra inclusion criteria for participants in the RCTs were: aged between 18 and 70 years, and having a positive (≥50% pain reduction 30-90 minutes after procedure) diagnostic facet joint or SI-joint block. Exclusion criteria for the RCTs were pregnancy, anticoagulant drug therapy and/or coagulopathy, body mass index (BMI) >35, involvement in a work-related conflict, and severe psychiatric or psychological problems. More details on the eligibility criteria are reported in the study protocol.[7]

For the current study, in which we compare the randomised study groups to the observational study groups, we selected the following groups:

- Patients with CLBP originating from the *facet joints*, receiving RF denervation and an exercise programme
    1. Randomised study group (intervention group of the RCT)
    2. Observational study group, matched to the randomised group fulfilling the RCT in- and exclusion criteria
    3. Observational study group, patients who were not matched to the RCT based on in- and exclusion criteria; e.g., they were older than 70 years, had a BMI >35 and/or psychological problems

- Patients with CLBP originating from the *SI-joints,* receiving RF denervation and an exercise programme
    1. Randomised study group (intervention group of the RCT)
    2. Observational study group, matched to the randomised group fulfilling the RCT in- and exclusion criteria
    3. Observational study group, patients who were not matched to the RCT based on in- and exclusion criteria; e.g., they were older than 70 years, had a BMI >35 and/or psychological problems

Study interventions

Each participant in the RCT received a standardised exercise programme of three months (range in total between eight and 12 hours) combined with psychological support if necessary. Participants in the intervention group also received RF denervation. This included facet joint RF denervation, or Cooled RF denervation, Simplicity III probe or Palissade technique as treatment for SI-joint pain.[3,8,9] Participants were asked to refrain from any co-intervention during the three-month intervention period. Anaesthesiologists 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 standardised exercise programme. The psychological interventions, if necessary, took place in a primary care setting.

Participants in the observational study did not receive a standardised exercise programme, but were only monitored prospectively. For the current study, only participants who received facet joint or SI-joint RF denervation and an exercise programme were selected out of the entire observational cohort.

Outcomes

The three primary outcome measures were pain intensity (11-point Numerical Rating Scale (NRS)),[10] functional status (Oswestry Disability Index 0-100 (ODI)),[11] and recovery (global perceived effect (GPE), 7-point Likert scale).[12] Treatment success on the GPE was achieved if a patient reported being “much recovery” or “complete recovery”.

Patients in the MinT study were followed up for 12 months. For the current study, only the results of the three months follow up are used, because this was the primary outcome assessment and we expected that the effect of randomisation would only play a role during the intervention period. All questionnaires were web-based and sent at

baseline and three months after the start of the treatment. Pain intensity and recovery were assessed at three and six weeks after start of treatment as well.

Statistical methods

Baseline characteristics of participants in the intervention group of the RCTs and in the observational study were compared using descriptive statistics. We compared each of the two observational study groups pairwise with the randomised study groups for CLBP originating from the facet joints and SI joints separately. Baseline characteristics were compared between completers and non-completers to identify possible selective dropout.

We adjusted for the possible effect of missing data on the study results in the analysis of mean changes for continuous outcomes by using maximum likelihood estimation for longitudinal mixed-effects models under "missing at random" assumptions and including a term for pain clinic if necessary.[13] We used a generalized linear mixed model (logit link) with the same multilevel structure for dichotomous outcomes. We calculated regression coefficients or odds ratios with 95% confidence intervals (CI). We adjusted for baseline characteristics related to outcome, and were interested in the time by treatment interaction. We used MLwiN to analyse the data (V2.22), with a level of significance of $p<.05$.

RESULTS

In total, 7529 participants were included in one of the RCTs or observational study of the MinT study between January 1, 2013 and December 17, 2015 and assessed for eligibility in this analysis. Patients were excluded from the current study if they participated in the control group of the RCTs, were not treated with RF denervation, did not have isolated facet joint or SI-joint pain, or if inclusion criteria or treatment details were unknown. All reasons for exclusion from the RCT are presented in Appendix 1. Most people were excluded from the RCTs because of psychological problems and age >70 years.

In total, 1421 participants fulfilled the selection criteria for the current study (see Figure 1).

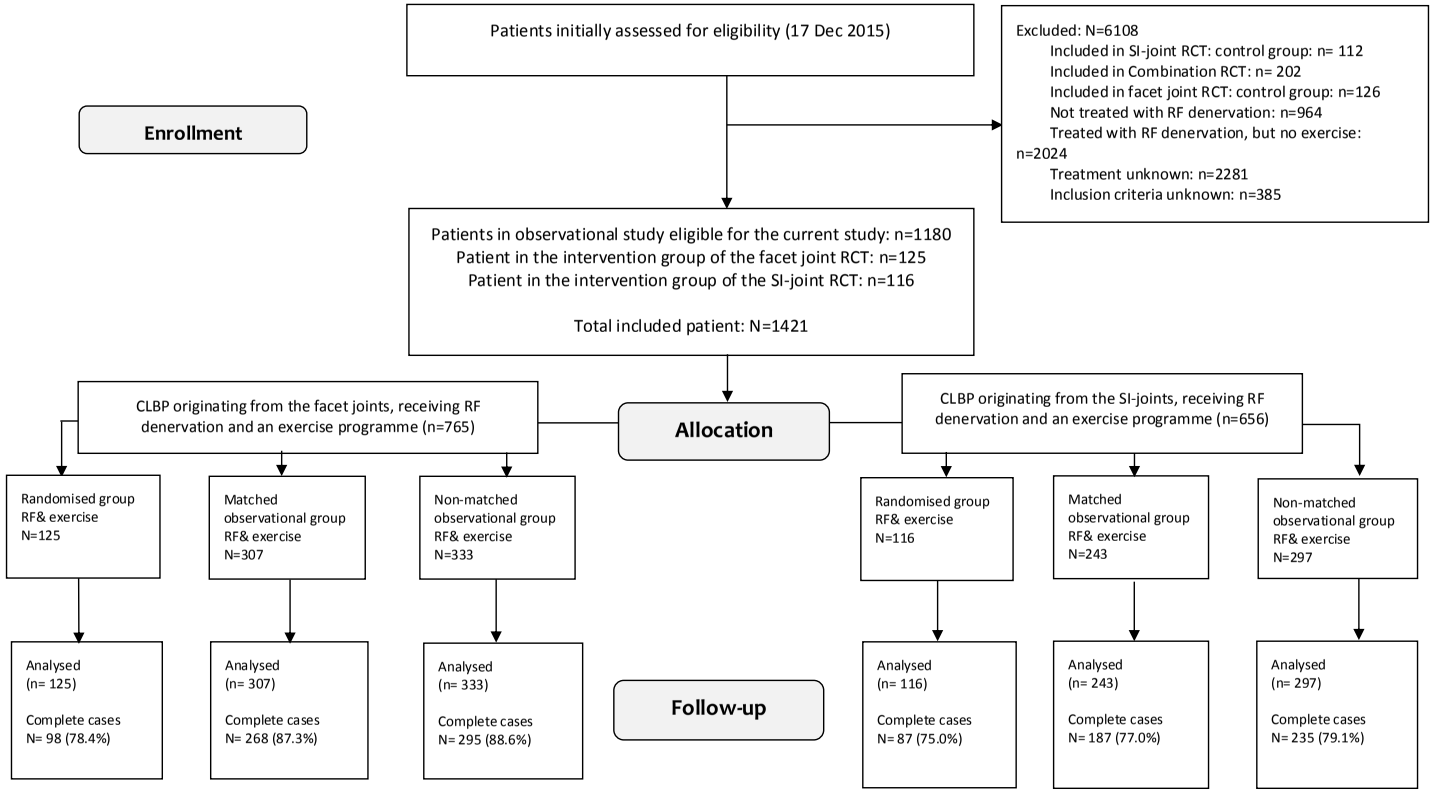


Figure 1. Flow chart

From the patients with LBP originating from the facet joints, 125 patients were randomised in the intervention group, 307 patients participated in the matched observational study group and 297 patients participated in the non-matched observational study group. From the patients with LBP originating from the SI-joints, 116 patients participated in the randomised study group, 243 patients in the matched and 297 in the non-matched observational study group.

Patient characteristics

Baseline characteristics are shown in Table 1. The total study population had a mean age of 55.6 years (SD 13.4) the majority were women, the population was mostly low educated, married, and had on average LBP complaints for 13 years (SD 12.5). During the three months, the majority of the participants sought primary healthcare more than 10 times, visited the outpatient-clinic at least once, and were never hospitalised (Table 2). On average 20% of the participants used weak opioids, and 4.3%-15.2% used strong opioids (Table 2).

Overall, all groups were similar, but participants in the non-matched observational study groups were older, less educated, less likely to have a paid job, had more functional limitations at baseline, and used more strong opioids compared to the participants in the randomised and matched observational study groups (Table 1 and 2). This applied to participants with CLBP originating from the facet joints as well as the SI-joints.

Within the patients with facet joint pain and SI-joint pain, there were no differences between completers and non-completers in baseline characteristics, healthcare use and medication, and outcomes (Appendix 2 and Appendix 3). The number of complete cases ranged between the groups from 75% to 88% (Figure 1).

Table 1. Descriptive baseline characteristics – CLBP originating from the facet joints and the SI-joints

	Participants with CLBP originating from the facet joints*			Participants with CLBP originating from the SI-joints*		
Characteristics	Randomised group N=125	Matched observational group N=307	Non-matched observational group N=333	Randomised group N=116	Matched observational group N=243	Non-matched observational group N=297
Age in years (SD)	52.9 (11.5)	52.8 (11.2)	61.8 (13.7)	51.6 (10.9)	51.2 (10.9)	57.6 (15.3)
Female (N (%))	65 (55.6%)	195 (63.5%)	310 (63.1%)	87 (75.0%)	185 (79.7%)	230 (80.7%)
BMI (SD)	26.7 (5.2)	26.9 (3.9)	28.5 (5.5)	26.7 (4.2)	26.6 (3.8)	28.9 (6.2)
Smoker (N (%))	34 (29.1%)	80 (26.1%)	76 (22.8%)	29 (26.6%)	54 (23.3%)	64 (22.5%)
Education						
• Low (N (%))	57 (48.7%)	132 (43%)	190 (57.1%)	59 (54.1%)	97 (41.8%)	159 (53.5%)
• Moderate (N (%))	35 (29.9%)	100 (32.6%)	89 (26.7%)	32 (29.4%)	90 (37.0%)	82 (29.0%)
• High (N (%))	21 (17.9%)	75 (24.4%)	54 (16.2%)	18 (16.5%)	45 (18.5%)	42 (14.8%)
History of back pain complaints						
• Months first LBP experience (median (IQR))	146 (50-267)	122 (39-243)	123 (49-247)	97 (37 -228)	121 (42 – 219)	103 (36 – 243)
• Months with current LBP episode (median (IQR))	31 (12-103)	28 (12-62)	40 (12-118)	30 (12 – 76)	30 (12 – 85)	30 (12 – 97)
Married/living with a partner (N (%))	93 (74.4%)	249 (71.1%)	239 (71.8%)	85 (78.0%)	178 (76.7%)	197 (69.4%)
Expectations						
• Credibility (0-27)	21.4 (3.9)	21.4 (4.0)	21.4 (4.2)	21.4 (4.5)	22.4 (3.4)	21.2 (4.0)
• Expectancy (0-27)	18.9 (4.6)	18.6 (4.6)	18.1 (4.7)	18.8 (4.9)	19.5 (4.2)	18.2 (4.8)
Having a paid job	64 (51.2%)	173 (56.4%)	81 (24.3%)	66 (61.7%)	137 (56.4%)	98 (33.4%)

Abbreviations: SD, Standard Deviation; N, number; BMI, Body Mass Index; EQ-5D, EuroQol-5D;
** Results are presented of the participants who had complete baseline data*

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

	Participants with CLBP originating from the facet joints			Participants with CLBP originating from the SI-joints		
Characteristics	**Randomised group N=125**	**Matched observational group N=307**	**Non-matched observational group N=333**	**Randomised group N=116**	**Matched observational group N=243**	**Non-matched observational group N=297**
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%)	130 (42.3%)	146 (43.8%)	32 (29.4%)	108 (44.4%)	129 (43.4%)
• ≥10 (%)	80 (67.2%)	177 (57.7%)	187 (56.2%)	77 (70.6%)	135 (55.6%)	168 (56.6%)
Outpatient clinic						
• 0 (%)	49 (41.2%)	92 (30.0%)	97 (29.1%)	43 (39.4%)	89 (36.6%)	100 (33.7%)
• ≥1 (%)	70 (58.8%)	215 (70.0%)	236 (70.9%)	66 (56.9%)	154 (63.4%)	197 (66.3%)
One day treatment						
• 0 (%)	77 (64.7%)	171 (55.7%)	203 (61.0%)	62 (56.9%)	141 (58.0%)	178 (59.9%)
• ≥1 (%)	42 (35.3%)	136 (44.3%)	130 (39.0%)	47 (43.1%)	102 (42.0%)	119 (40.1%)
Hospitalisation						
• 0 (%)	119 (100.0%)	301 (98.0%)	328 (98.5%)	108 (99.1%)	237 (97.5%)	292 (98.3%)
• ≥1 (%)	0 (0.0%)	6 (2.0%)	5 (1.5%)	1 (0.9%)	6 (2.5%)	5 (1.7%)
Medication use						
• None/non back pain related	46 (39.7%)	91 (29.8%)	102 (31.1%)	37 (34.3%)	65 (26.9%)	77 (26.3%)
• Non-opioids (%) (aspirin/paracetamol/ NSAID)	42 (36.2%)	126 (41.3%)	111 (33.7%)	53 (38.9%)	99 (40.9%)	110 (37.5%)
• Weak opioids (%) (with or without non-opioids)	23 (19.8%)	56 (18.4%)	66 (20.1%)	24 (22.2%)	55 (22.7%)	68 (23.2%)
• Strong opioids (%) (with or without non-opioids)	5 (4.3%)	32 (10.5%)	50 (15.2%)	5 (4.6%)	23 (9.5%)	38 (13.0%)

Comparison between the randomised study groups and the observational study groups

Facet joints

For all participants with CLBP originating from the facet joints, the mean pain intensity decreased 1.97 points in the randomised study group in the first three weeks after receiving RF denervation, 2.03 points in the matched observational group, and 1.68 in the non-matched observational study group. Pain intensity stabilised afterwards (Table 3 and Figure 2). There were no differences in the course of pain intensity and recovery for the matched and non-matched observational study groups when compared to the randomised study group (Table 3 and Figure 2). We found a statistically significant difference in functional status in where there was more improvement in the randomised study group compared to the non-matched observational study group (mean difference (MD) 4.34; 95% confidence interval (CI): 1.42-7.25) on a 0-100 scale, but no statistical significant differences when compared to the matched observational study group.

SI-joints

For all participants with CLBP originating from the SI-joints, the mean pain intensity decreased 2.21 points in the randomised study group in the first three weeks after receiving RF denervation, 2.32 points in the matched observational group, and 1.96 in the non-matched observational study group. This stabilised afterwards (Table 3 and Figure 2). We found no differences in course of LBP symptoms between the matched observational study groups and the randomised study group. Participants in the randomised study group had statistically significant more pain reduction (MD 0.46; 95%CI: 0.04-0.88) on a 0-10 scale and more improvement in functional status (MD 6.41; 95%CI: 3.30-9.51) on a 0-100 scale compared to participants in the non-matched observational study group over the three month period.

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

		Participants with CLBP originating from the facet joints					Participants with CLBP originating from the SI-joints				
		Randomised group N=125	**Matched observational group N=307**	**Mean difference (95%CI)**	**Non-matched observational group N=333**	**Mean difference (95%CI)**	**Randomised group N=116**	**Matched observational group N=243**	**Mean difference (95%CI)**	**Non-matched observational group N=297**	**Mean difference (95%CI)**
NRS Pain (SD)	Overall effect			-0.12 (-0.56-0.18)		0.08 (-0.31-0.46)			0.08 (-0.35-0.50)		0.46 (0.04-0.88)*
	Baseline	7.14 (1.38)	7.06 (1.46)		7.44 (1.52)		7.17 (1.65)	7.55 (1.34)		7.38 (1.47)	
	3 weeks	5.17 (2.27)	5.03 (2.15)	-0.32 (-0.79-0.15)	5.76 (2.18)	0.12 (-0.35-0.59)	4.96 (2.19)	5.23 (2.19)	0.18 (-0.36-0.72)	5.42 (2.28)	0.44 (-0.10-0.97)
	6 weeks	5.19 (2.31)	5.04 (2.23)	-0.24 (-0.69-0.21)	5.49 (2.06)	-0.07 (-0.53-0.39)	5.22 (2.16)	4.85 (2.23)	-0.36 (-0.88-0.15)	5.62 (2.24)	0.31 (-0.20-0.81)
	3 months	5.01 (2.29)	5.01 (2.34)	-0.04 (-0.48-0.41)	5.53 (2.20)	0.17 (-0.28-0.63)	4.77 (2.46)	5.19 (2.39)	0.44 (-0.08-0.95)	5.45 (2.27)	0.64 (0.13-1.15)*
ODI Functioning (SD)	Baseline	35.08 (14.66)	38.03 (13.95)		44.52 (15.21)		38.07 (14.07)	41.45 (13.50)		45.17 (13.92)	
	3 months	26.03 (16.58)	29.95 (16.21)	1.44 (-1.37-4.24)	38.43 (17.37)	4.34 (1.42-7.25)**	27.72 (17.05)	31.61 (15.07)	2.83 (-0.26-5.92)	39.35 (18.54)	6.41 (3.30-9.51)***
				OR (95%CI)		**OR (95%CI)**			**OR (95%CI)**		**OR (95%CI)**
GPE recovery(%)	Overall effect			1.07 (0.73-1.59)		0.94 (0.62-1.41)			0.81 (0.54-1.24)		0.69 (0.46-1.06)
	3 weeks	32 (29.6%)	72 (26.6%)	0.93 (0.51-1.68)	68 (22.7%)	0.88 (0.48-1.60)	28 (29.8%)	58 (29.9%)	0.87 (0.46-1.65)	61 (24.7%)	0.76 (0.40-1.43)
	6 weeks	35 (29.4%)	97 (32.3%)	1.27 (0.73-2.23)	92 (28.4%)	1.20 (1.34-2.12)	40 (37.0%)	80 (34.2%)	0.77 (0.43-1.37)	80 (27.8%)	0.71 (0.40-1.25)
	3 months	43 (36.1%)	108 (35.2%)	1.03 (0.60-1.76)	88 (26.4%)	0.78 (0.45-1.36)	43 (39.1%)	93 (38.3%)	0.82 (0.46-1.44)	86 (29.0%)	0.64 (0.36-1.14)

Values presented are model estimates of linear mixed-effects models with a random intercept, and adjusted for outcomes at baseline and age, gender, BMI, education, smoking, marital status, back pain complaint history, patient expectations and medication use. 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. P<0.05; **P<0.01; *** P<0.001*

Figure 2.1. Patients with CLBP originating from the facet joints

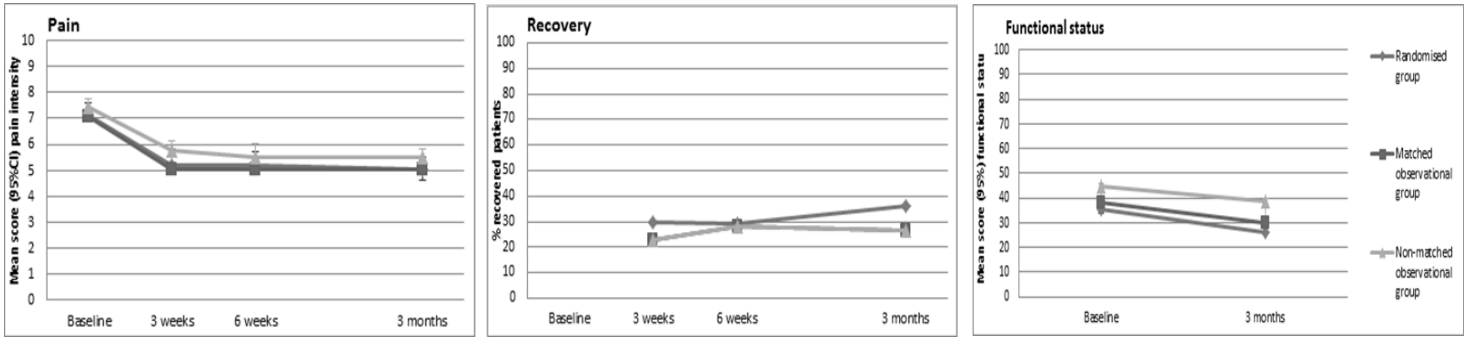
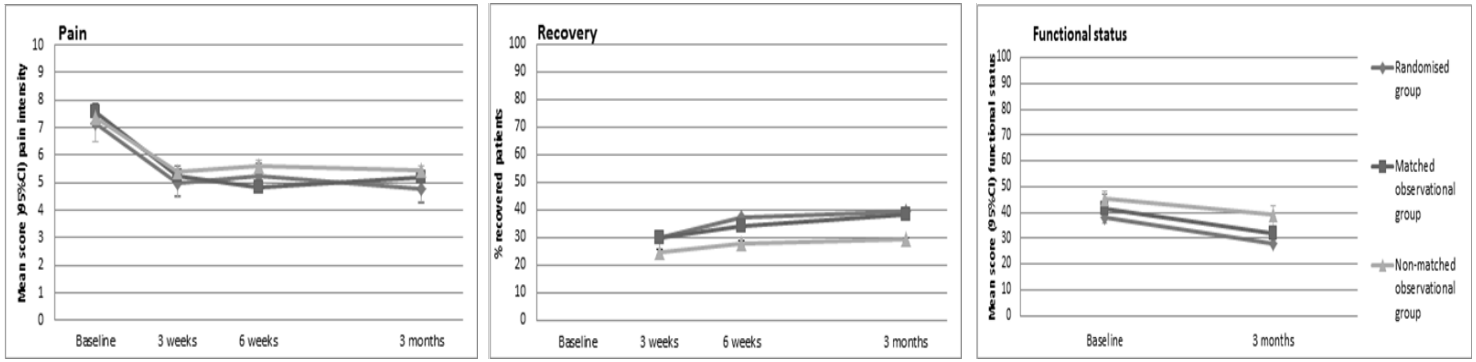


Figure 2.2. Patients with CLBP originating from the SI-joints



Sensitivity analysis

Due to the selection of participants, there was no missing data at the three months assessment in the observational study groups. In the observational facet joints groups, there were almost 10% more complete cases compared to the randomised study groups (78.4% versus respectively 87.3% and 88.6%). For participants with CLBP originating from the SI-joints, the number of complete cases was 75.0% in the randomised study group and 77.0% and 79.0% in the matched and non-matched observational study groups respectively (see Figure 1).

Analysing only complete cases slightly reduced the mean differences between the groups, showing no statistically significant differences in pain intensity between the non-matched observational study group and the randomised study group three months after the RF denervation (MD 6.03; 95%CI 2.64-9.43) (Appendix 4 and 5). No other differences between the main analysis and complete case analysis were found.

Models without adjustments for baseline differences (data not shown) showed larger mean differences between the non-matched observational study groups and the randomised study groups in favour of the randomised study group. This was to be expected from regression to the mean, since the non-matched observational study groups started out with worse pain intensity and more limitations in functional status.

DISCUSSION

The current study compared the course in symptoms between randomised study groups and observational study groups for patients with CLBP originating from the facet joints or SI-joints who were treated with RF denervation and an exercise programme. All study groups showed small improvements over time, and all study groups showed a very comparable course on all outcomes. For participants in the matched observational study groups, we found no differences on any outcome compared to the randomised study groups. Participants in the non-matched observational study groups with facet joint or SI-joint pain showed slightly more limitations in functional status three months after the RF denervation compared to the participants in the randomised study groups. Participants with SI-joint pain in the non-matched observational study group also showed statistically significant higher pain intensity at the three months assessment compared to participants in the randomised study group. The differences between the non-matched observational study groups and the randomised study groups are statistically significant, but small. Previous studies estimated minimal important change scores for patients with LBP of eight to 12 points on the ODI.[14,15] All study groups in the MinT study showed absolute changes in functional status over time on the ODI of less than 10 points, and the adjusted difference in functional status between the groups was less than 10 points as well. The clinical relevance of the differences between the groups in the current study can be questioned.

A strength of the current study is the nationwide study design, which resulted in a large sample of patients that are recruited in routine clinical care. This increases the applicability of the RCT results and the observational study results in daily practice. Secondly, patients who were randomly allocated to the intervention could be compared to patients who had a treatment choice in the observational study group. As we found no clinically relevant differences between these groups, our results challenge the assumption that the course of symptoms as identified in randomised groups are not straightforward applicable to situations beyond the scope of the randomised study setting. Or in other words, it shows that the course of their symptoms is probably not influenced due to their participation in the randomised trials. Therefore, the current study provides evidence for external validation of RCT results and supports the generalizability of the results.

A possible limitation of the current study is the fact that other patients may present themselves at the pain clinic and are less comparable to the patients we included in

the current study. However, the inclusion criteria of the RCTs were defined in close consultation with participating physicians in our study and correspond to the inclusion criteria for treatment in daily practice. For that reason, we expect that the patients in this study are a good reflection of the patients in daily practice. Secondly, a limitation of this study was the inability to select a control group in the observational data. Therefore, it was not possible to perform a non-randomised comparison of RF denervation in addition to an exercise programme versus an exercise programme alone. Various studies and one meta-analysis did perform a non-randomised treatment comparison in the field of LBP.[16-19] These studies showed similar results in clinical course of LBP symptoms compared to the current study. However, these previous studies analysed a variety of mostly conservative treatments, and clinical and statistical heterogeneity could potentially have influenced the results. The current study adds to the existing literature in the sense that it illustrates that course of symptoms as found in the RCTs in a secondary care setting for patients who received a RF denervation at the pain clinic are comparable to daily practice at the pain clinic.

Current systematic reviews and meta-analyses often ignore the evidence generated in observational studies. Reasons for this are the lack of a control group and the assumption that their findings are more likely to be biased. Usually, the difference in results between RCTs and observational studies are attributed to differences in methodological quality. A meta-analysis by Furlan et al undermines this assumption.[1] Furlan et al. compared clinical outcomes of RCTs and non-randomised studies in the field of LBP regarding their methodological quality and heterogeneity, and showed mostly similar results.[1] The results of the study of Furlan increase the body of evidence for the comparability between RCT results and observational study results in the field of LBP.

For a long time, it was believed that observational studies overestimate the effects of RCTs.[20] But more data has become available that observational studies show similar results as RCTs.[2,21] Our study results are in line with these findings, and it is more likely that differences between studies are attributed to study setting, population and intervention, and not to study design itself.[22]

Conclusion

Despite the belief that observational studies might overestimate the effects of RCTs, our study showed a largely similar course of LBP symptoms over a period of three months in patients with CLBP originating from the facet joints and SI-joints in a randomised study

group compared to patients in a observational study group. Results from the RCTs seem to be generalizable to a similar patient population in clinical practice, and only small differences were shown when comparing RCT results to results in an observational study group with a slightly different population.

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 participants, physiotherapists, and everyone working on the MinT study at one of the participating pain clinics.

REFERENCES

1. 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.
2. Artus M, van der Windt D, Jordan KP, Croft PR. The clinical course of low back pain: a meta-analysis comparing outcomes in randomised clinical trials (RCTs) and observational studies. *BMC Musculoskelet Disord.* 2014;15(1):68.
3. Cosman Jr ER, Gonzalez CD. Bipolar radiofrequency lesion geometry: implications for palisade treatment of sacroiliac joint pain. *Pain Practice.* 2011;11(1):3-22.
4. Kline M. Radiofrequency techniques in clinical practice. *Interventional Pain Management.* 2000.
5. Cohen S, Huang J, Brummett C. Facet joint pain—advances in patient selection and treatment. *Nat R Rheumatol.* 2012;9:101-116.
6. Cohen SP, Chen Y, Neufeld NJ. Sacroiliac joint pain: a comprehensive review of epidemiology, diagnosis and treatment. *Expert Rev Neurother*. 2013; 13(1):99-116.
7. Maas E, Juch J, Groeneweg J, 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(1):260.
8. Cohen SP, Hurley RW, Buckenmaier III CC, Kurihara C, Morlando B, Dragovich A. Randomized placebo-controlled study evaluating lateral branch radiofrequency denervation for sacroiliac joint pain. *Anesthesiology.* 2008;109(2):279.
9. Schmidt PC, Pino CA, Vorenkamp KE. Sacroiliac joint radiofrequency ablation with a multilesion probe: A case series of 60 patients. *Anestha & Analg.* 2014;119(2):460-462.
10. Downie W, Leatham P, Rhind V, Wright V, Branco J, Anderson J. Studies with pain rating scales. *Annals of The Rheumatic Diseases.* 1978;37(4):378-381.
11. Fairbank J, Park W, McCall I, O'Brien J. Apophyseal injection of local anesthetic as a diagnostic aid in primary low-back pain syndromes. *Spine.* 1981;6:598-605.
12. Kamper SJ, Maher CG, Herbert RD, Hancock MJ, Hush JM, Smeets RJ. How little pain and disability do patients with low back pain have to experience to feel that they have recovered? *Eur Spine J.* 2010;19(9):1495-1501.
13. Twisk J. *Applied multilevel analysis: a practical guide for medical researchers.* Cambridge Univ Press; 2006.
14. Ostelo RW, de Vet HC. Clinically important outcomes in low back pain. *Best Prac Res Clin Rheumatol.* 2005;19(4):593-607.
15. Hägg O, Fritzell P, Nordwall A. The clinical importance of changes in outcome scores after treatment for chronic low back pain. *Eur Spine J.* 2003;12(1):12-20.
16. Weinstein JN, Lurie JD, Tosteson TD, et al. Surgical versus non-operative treatment for lumbar disc herniation: four-year results for the Spine Patient Outcomes Research Trial (SPORT). *Spine.* 2008;33(25):2789.
17. 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.
18. Bonetti F, Curti S, Mattioli S, et al. Effectiveness of a'Global Postural Reeducation'program for persistent low back pain: a non-randomized controlled trial. *BMC Musculoskelet Disord.* 2010;11(1):285.

19. 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.
20. Kunz R, Oxman AD. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. *BMJ.* 1998;317(7167):1185-1190.
21. Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. *NEJM.* 2000;342(25):1878-1886.
22. Shrier I, Boivin J-F, Steele RJ, 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.* 2007;166(10):1203-1209.

Appendix 1. Reasons for exclusion from the RCT

Exclusion criteria	Participants with CLBP originating from the facet Joints RCT N=333	Participants with CLBP originating from the SI-joints RCT N=297
Psychological problems	139 (41.7%)	145 (48.8%)
Age >70	77 (23.1%)	40 (13.5%)
Psychological problems & age>70	29 (8.7%)	20 (6.7%)
Negative diagnostic block	23 (6.9%)	28 (9.4%)
BMI>35	22 (6.6%)	25 (8.4%)
Psychological problems & negative diagnostic block	16 (4.8%)	15 (5.1%)
Psychological problems & BMI>35	15 (4.5%)	13 (4.4%)
Negative diagnostic block & age>70	5 (1.5%)	2 (0.7%)
Age>70 & BMI>35	2 (0.6%)	3 (1.0%)
Psychological problems & age>70 & BMI>35	1 (0.3%)	3 (1.0%)
Negative diagnostic block & BMI>35	2 (0.6%)	0 (0.0%)
Psychological problems & BMI>35 & negative diagnostic block	2 (0.6%)	1 (0.3%)
Psychological problems & age>70 & negative diagnostic block	0 (0.0%)	2 (0.7%)
Age>70 & BMI>35 & negative diagnostic block	0 (0.0%)	0 (0.0%)

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

	Participants with CLBP originating from the facet joints			Participants with CLBP originating from the SI-joints		
Characteristics	**Randomised group N=98**	**Matched observational group N=268**	**Non-matched observational group N=295**	**Randomised group N=87**	**Matched observational group N=187**	**Non-matched observational group N=235**
Age in years (SD)	53.6 (11.3)	53.0 (11.2)	61.9 (13.8)	52.9 (10.9)	50.9 (11.1)	57.7 (15.3)
Female (N (%))	54 (55.1%)	170 (63.4%)	186 (63.1%)	70 (80.5%)	151 (80.7%)	185 (78.7%)
BMI (SD)	26.6 (5.2)	26.9 (3.9)	28.6 (5.5)	26.5 (4.2)	26.5 (3.8)	29.0 (6.4)
Smoker (N (%))	29 (29.6%)	70 (26.1%)	68 (23.1%)	25 (28.7%)	43 (23.0%)	56 (23.8%)
Education						
• Low **(N (%))**	48 (49.0%)	121 (45.1%)	172 (58.3%)	49 (56.3%)	79 (42.2%)	139 (55.1%)
• Moderate **(N (%))**	29 (29.6%)	84 (31.3%)	77 (26.1%)	25 (28.7%)	73 (39.0%)	68 (29.1%)
• High **(N (%))**	21 (21.4%)	63 (23.5%)	46 (15.6%)	13 (14.9%)	35 (18.7%)	37 (15.8%)
History of back pain complaints						
• Months first LBP experience (median (IQR))	146 (52-271)	122 (37-243)	146 (49-272)	120 (37-231)	113 (42-198)	107 (37-243)
• Months with current LBP episode (median (IQR))	30 (12-97)	26 (12-61)	43 (12-122)	30 (10-79)	30 (12-85)	30 (12-73)
Married/living with a partner (N (%))	77 (78.6%)	218 (81.3%)	210 (71.2%)	70 (80.5%)	143 (76.5%)	167 (71.1%)
Expectations						
• Credibility (0-27)	21.5 (2.59)	21.4 (4.0)	21.3 (4.3)	21.1 (4.5)	22.6 (3.3)	21.3 (3.9)
• Expectancy (0-27)	19.2 (4.4)	18.5 (4.7)	18.5 (5.1)	18.5 (4.9)	19.6 (4.3)	18.3 (4.7)
Having a paid job	52 (53.6%)	154 (57.7%)	70 (23.7%)	51 (60.0%)	110 (58.8%)	81 (34.8%)

Abbreviations: SD, Standard Deviation; N, number; BMI, Body Mass Index

Appendix 3. Complete case descriptive healthcare and medication use – CLBP originating from the facet joints and the SI-joints

	Participants with CLBP originating from the facet joints			Participants with CLBP originating from the SI-joints		
Characteristics	**Randomised group N=98**	**Matched observational group N=268**	**Non-matched observational group N=295**	**Randomised group N=87**	**Matched observational group N=187**	**Non-matched observational group N=235**
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%)	112 (42.2%)	129 (43.7%)	25 (28.7%)	81 (43.3%)	101 (43.0%)
• ≥10 (%)	69 (70.4%)	155 (57.8%)	166 (56.3%)	62 (71.3%)	106 (56.7%)	134 (57.0%)
Outpatient clinic						
• 0 (%)	40 (40.8%)	83 (31.0%)	85 (28.8%)	35 (40.2%)	65 (34.8%)	79 (33.6%)
• ≥1 (%)	58 (59.2%)	185 (69.0%)	210 (71.2%)	52 (59.8%)	122 (65.2%)	156 (66.4%)
One day treatment						
• 0 (%)	65 (66.3%)	144 (53.7%)	178 (60.3%)	47 (54.0%)	103 (55.1%)	143 (60.9%)
• ≥1 (%)	33 (33.7%)	124 (46.3%)	117 (39.7%)	40 (46.0%)	84 (44.9%)	92 (39.1%)
Hospitalisation						
• 0 (%)	98 (100%)	264 (98.5%)	290 (98.4%)	86 (98.9%)	183 (97.9%)	231 (98.3%)
• ≥1 (%)	0 (0.0%)	4 (1.5%)	5 (1.7%)	1 (1.1%)	4 (2.1%)	4 (1.7%)
Medication use						
• None/non back pain related	36 (36.7%)	77 (28.9%)	92 (31.6%)	32 (37.2%)	54 (28.9%)	62 (26.6%)
• Non-opioids (%) (aspirin/paracetamol/ NSAID)	38 (39.2%)	111 (41.7%)	94 (32.3%)	31 (36.0%0	75 (40.3%)	92 (39.5%)
• Weak opioids (%) (with or without non-opioids)	19 (19.6%)	48 (18.0%)	60 (20.6%)	19 (21.8%)	38 (20.3%)	53 (22.7%)
• Strong opioids (%) (with or without non-opioids)	4 (4.1%)	30 (11.3%)	45 (15.5%)	4 (4.7%)	19 (10.2%)	26 (11.2%)

Appendix 4. Complete case analysis: Treatment effects of observational study groups compared to the RCT

		Participants with CLBP originating from the facet joints					Participants with CLBP originating from the SI-joints				
		Randomised group N=98	**Matched observational group N=268**	**Mean difference (95%CI)**	**Non-matched observational group N=295**	**Mean difference (95%CI)**	**Randomised group N=87**	**Matched observational group N=187**	**Mean difference (95%CI)**	**Non-matched observational group N=235**	**B (95%CI) group**
NRS Pain (SD)	Overall effect			-0.19 (-0.58-0.21)		0.17 (-0.23-0.58)			0.05 (-0.42-0.52)		0.29 (-0.17-0.75)
	Baseline	7.10 (1.41)	7.09 (1.40)		7.45 (1.52)		7.03 (1.76)	7.63 (1.24)		7.37 (1.47)	
	3 weeks	5.20 (2.22)	5.03 (2.15)	-0.27 (-0.75-0.21)	5.74 (2.18)	0.19 (-0.31-0.68)	5.01 (2.21)	5.19 (2.21)	0.10 (-0.47-0.67)	5.43 (2.29)	0.25 (-0.30-0.81)
	6 weeks	5.12 (2.33)	5.03 (2.24)	-0.18 (-0.67-0.30)	5.47 (2.04)	0.01 (-0.48-0.50)	5.21 (2.17)	4.93 (2.17)	-0.31 (-0.88-0.26)	5.54 (2.29)	0.20 (-0.36-0.76)
	3 months	4.87 (2.25)	5.04 (2.33)	0.10 (-0.38-0.58)	5.51 (2.23)	0.32 (-0.17-0.81)	4.74 (2.54)	5.12 (2.38)	0.36 (-0.21-0.93)	5.27 (2.28)	0.41 (-0.15-0.97)
ODI Functioning (SD)	Baseline	35.33 (14.49)	38.16 (13.77)		44.65 (15.36)		38.51 (14.05)	41.21 (13.43)		44.68 (13.78)	
	3 months	26.22 (17.12)	30.31 (16.41)	1.42 (-0.64-4.47)	38.32 (17.55)	4.24 (1.07-7.41)***	27.10 (17.51)	30.61 (14.67)	2.85 (-0.57-6.27)	37.74 (17.23)	6.03 (2.64-9.43)***
				OR (95%CI)		**OR (95%CI)**			**OR (95%CI)**		**OR (95%CI)**
GPE recovery(%)	Overall effect			1.03 (0.68-1.57)		0.88 (0.57-1.35)			0.76 (0.48-1.21)		0.75 (0.47-1.19)
	3 weeks	30 (30.6%)	71 (26.5%)	0.88 (0.48-1.62)	68 (23.1%)	0.83 (0.45-1.53)	26 (29.9%)	58 (31.0%)	0.90 (0.46-1.74)	58 (24.7%)	0.83 (0.43-1.60)
	6 weeks	29 (29.6%)	87 (32.5%)	1.30 (0.71-2.36)	84 (28.5%)	1.18 (0.64-2.17)	34 (39.1%)	63 (33.7%)	0.66 (0.35-1.26)	69 (29.4%)	0.70 (0.37-1.31)
	3 months	37 (37.8%)	94 (35.1%)	0.96 (0.54-1.72)	78 (26.4%)	0.71 (0.39-1.28)	35 (40.2%)	70 (37.4%)	0.75 (0.40-1.41)	74 (31.5%)	0.74 (0.39-1.38)

Valus presented are model estimates of linear mixed-effects models with a random intercept, and adjusted for outcomes at baseline and age, gender, BMI, education, smoking, marital status, back pain complaint history, patient expectations and medication use. 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. P<0.05; **P<0.01; *** P<0.001*