

# Effectiveness of stratified treatment for back pain in Danish primary care: A randomized controlled trial

Lars Morsø<sup>1</sup>  | Kim Olsen Rose<sup>2</sup>  | Berit Schiøttz-Christensen<sup>4</sup>  | Gail Sowden<sup>3</sup> | Jens Søndergaard<sup>4</sup>  | David H. Christiansen<sup>5</sup> 

<sup>1</sup>Clinical Department, University of Southern Denmark, Odense, Denmark

<sup>2</sup>Department of Business and Economics, DaCHE, University of Southern Denmark, Odense, Denmark

<sup>3</sup>Keele University, Keele, England

<sup>4</sup>Research Unit of General Practice, University of Southern Denmark, Odense, Denmark

<sup>5</sup>Department of Occupational Medicine, Regional Hospital West Jutland, University Research Clinic, Herning, Denmark

## Correspondence

Lars Morsø, Clinical Department, University of Southern Denmark, Odense, Denmark.

Email: lars.morsoe@rsyd.dk

## Funding information

The Central Denmark Region, the Region of Southern Denmark, the Danish Rheumatism Association, the Association of Danish Physiotherapists and the General Practitioners Quality and Postgraduate Foundation funded this study. The funders did not have any influence on study design, data collection, management, analysis, interpretation or the publication of results.

## Abstract

**Background:** A randomized controlled trial (RCT) of stratified care demonstrated superior clinical outcomes and cost-effectiveness for low back pain (LBP) patients in UK primary care. This is the first study in Europe, outside of the original UK study, to investigate the clinical efficacy and cost-effectiveness of stratified care compared with current practice for patients with non-specific LBP.

**Methods:** The study was a two-armed RCT. Danish primary care patients with LBP were randomized to stratified care ( $n = 169$ ) or current practice ( $n = 164$ ). Primary outcomes at 3- and 12-months' follow-up were Roland Morris Disability Questionnaire (RMDQ), patient-reported global change and time off work. Secondary outcomes included pain intensity, patient satisfaction, healthcare resource utilization and quality-adjusted life years.

**Results:** Intention-to-treat analyses found no between-group difference in RMDQ scores at 3 months (0.5, 95% CI  $-1.8$  to  $0.9$ ) or 12 months (0.4,  $-2.1$  to  $1.3$ ). No overall differences were found between the arms at 3 and 12 months with respect to time off work or secondary outcomes. Stratified care intervention resulted in significantly fewer treatment sessions (3.5 [ $SD$  3.1] vs. 4.5 [3.5]) and significantly lower total healthcare costs (€) (13.4 [529] vs. 228 [830],  $p = .002$ ). There was no difference in cost-effectiveness (0.09, 0.05 to 0.13 vs. 0.10, 0.07–0.14,  $p = .70$ ).

**Conclusions:** There was no significant difference in clinical outcomes between patients with non-specific LBP receiving stratified care and those receiving current practice. However, stratified care may reduce total healthcare costs if implemented in Danish primary care.

**Significance:** Stratified care for low back pain based on risk profile is recommended by recent evidence based clinical guidelines. This study is the first broad replication of the STarT Back Trial in Europe. Therefore, the study adds to the body of knowledge evaluating the effectiveness of stratified care for low back pain in primary care, and provides insight into the effects of stratification on clinical practice.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2021 The Authors. *European Journal of Pain* published by John Wiley & Sons Ltd on behalf of European Pain Federation - EFIC ®

## 1 | INTRODUCTION

The concept of stratified care for low back pain (LBP) has gained popularity over recent decades (Flynn et al., 2002; Foster et al., 2013; Fritz & George, 2000). There are several theories and methods of subgrouping and stratifying people with LBP (Foster et al., 2011; Freynhagen et al., 2006; Fritz et al., 2007; Hartvigsen et al., 2018; Hodges, 2019; Konstantinou et al., 2018; Molgaard Nielsen et al., 2017; Oliveira, 2019). The STarT Back approach to stratified care for LBP was developed in 2008 (Hill, 2008) and was subsequently evaluated in two United Kingdom (UK) studies (Foster et al., 2014; Hill et al., 2011). The approach involves allocating LBP patients to one of three predefined subgroups (low-, medium- and high-risk subgroups) according to their risk of a poor outcome, based on the brief self-reported STarT Back Tool (SBT) questionnaire. Patients then receive the appropriate matched treatment for that subgroup (Hill, 2008).

The UK STarT Back trial (Hill et al., 2011) compared stratified care with best practice, and demonstrated that stratified care was superior in terms of improved clinical outcomes, reduced costs and increased efficiency of healthcare delivery. Following the STarT Back trial, the IMPaCT study (Foster et al., 2014) tested whether stratified care could be implemented into routine clinical care and the impact of doing so on physician clinical behaviour, patient outcomes and costs. The IMPaCT study was a prospective population-based sequential comparison study (Sowden et al., 2012). Results showed that stratified care resulted in significant improvement in patient disability outcomes and a 50% reduction in time off work without increased healthcare costs (Foster et al., 2014).

Despite the importance of testing stratified care for LBP in different healthcare contexts, few studies have done this, and those published have had different methodological designs (Cherkin et al., 2016, 2018; Riis et al., 2016; Riis et al., 2017; Werneke et al., 2020). The MATCH study (Cherkin et al., 2016) was the first study to evaluate the implementation of the stratified care approach outside the UK. However, the matched treatments in this trial (Cherkin et al., 2016) were different from those in the STarT Back trial. In the STarT Back trial intervention arm, the high-risk-matched treatment consisted of psychologically informed physiotherapy (PIP) (Main et al., 2012). However, within the MATCH trial, PIP was just one of several treatment options available to physicians for high-risk patients (Main et al., 2012). In the MATCH study patient outcomes and healthcare utilization was no different in the implementation arm compared with the usual care arm (Cherkin et al., 2018). There may be important differences between different healthcare contexts, however, no study has tried to replicate the original STarT Back trial, in Europe, outside of the original UK study.

This study, replicated the STarT Back trial with some adaptation to suit the Danish primary healthcare setting, and the Southern and Central Denmark evidence-based guideline recommendations. The aim of this study was to evaluate the clinical efficacy and cost-effectiveness of stratified care compared with current practice for patients with non-specific LBP in Danish primary care.

## 2 | METHODS

### 2.1 | Design and setting

The study protocol has been published (Morso et al., 2018). In brief, we performed a two-armed parallel-randomized controlled trial with a 1:1 allocation ratio to either stratified care (intervention) or current care (control) (see Appendix 1). The study was conducted in Danish primary healthcare in two of the five regions of Denmark, with a combined population size of 2.1 million citizens.

We invited general practitioners (GPs) and physiotherapists from 10 different geographical areas in the Regions of Southern and Central Denmark to take part in this study. In total, 42 general practices (102 GPs) and 74 physiotherapists from 21 physiotherapy clinics (two from each city, and three in one city) attended one of a number of local information meetings and subsequently agreed to participate in the study. Prior to patient inclusion half of the physiotherapy clinics (11 clinics) in the 10 cities, were randomly selected to deliver stratified care, and the other half (10 clinics) to deliver current practice (Morso et al., 2018). The two clinics in each city matched each other in size and employees. The patients in each city were randomized to either stratified care or the usual care control arm, and attended the clinic delivering that arm of the study.

### 2.2 | Study participants and recruitment procedure

The GPs were asked to identify and refer patients with LBP to primary care physiotherapy clinics. Simultaneously, they notified the project secretary, who contacted the patient by phone, checked eligibility, answered any questions and discussed potential involvement in the study. If the patient agreed to participate, the project secretary sent out electronic consent forms and baseline questionnaires including the SBT (Morso et al., 2018) to potential study participants. Following completion of the consent and baseline questionnaires, a database algorithm randomized participants. Specially trained physiotherapists provided the appropriate low-, medium- and high-risk matched treatments for participants in the stratified care arm of the study. Participants were recruited

to the study between December 2015 and December 2017, and the last 12-month follow-up questionnaire was completed on 5 December 2018. Because of slow recruitment, we extended the recruitment period, and in addition to normal recruitment, we allowed participating physiotherapists to recruit referred patients who were initially missed at the GP into the study. The Danish ethical committee approved this change (Project-ID: S-20140205), and we amended the clinical trial registration accordingly. Design, data collection and randomization procedures remained the same throughout the study.

### 2.3 | Inclusion and exclusion criteria

The study inclusion criteria were adults (18 years and over) with non-specific LBP (with or without associated leg pain) who were suitable for referral to physiotherapy by the GP and who were able to understand verbal and written Danish language. Exclusion criteria were serious or potentially serious pathology, serious illness or influential comorbidity, psychiatric illness, spinal surgery during the last 6 months, pregnancy or currently receiving physiotherapy for LBP (Morso et al., 2018).

### 2.4 | Randomization and blinding

After receiving participants' signed consent forms and completed baseline questionnaires, the central project secretary randomized participants at individual level into one of the two treatment arms, stratified by city and SBT subgroup, using computer-generated random number sequencing (see flow chart, Figure 1). After randomization, the project secretary contacted the relevant physiotherapy clinic to ensure initiation of treatment. Neither the treating physiotherapists nor the investigators had an influence on the allocation of participants in the study. Follow-up data collection time points were at 3 (primary clinical outcome point) and 12 months (time point for cost-effectiveness analysis). All questionnaires and reminders to complete questionnaires were sent electronically (text and mail) to patients. Randomization and data collection were administrated by an online clinical database specifically developed for this study.

### 2.5 | Stratified care (intervention arm)

To broadly replicate the STarT Back trial, we provided similar matched treatments for participants in the stratified care arm of the study as those provided in the STarT Back trial (Hill et al., 2011; Main et al., 2012). In addition, all stratified care arm participants received a patient information leaflet,

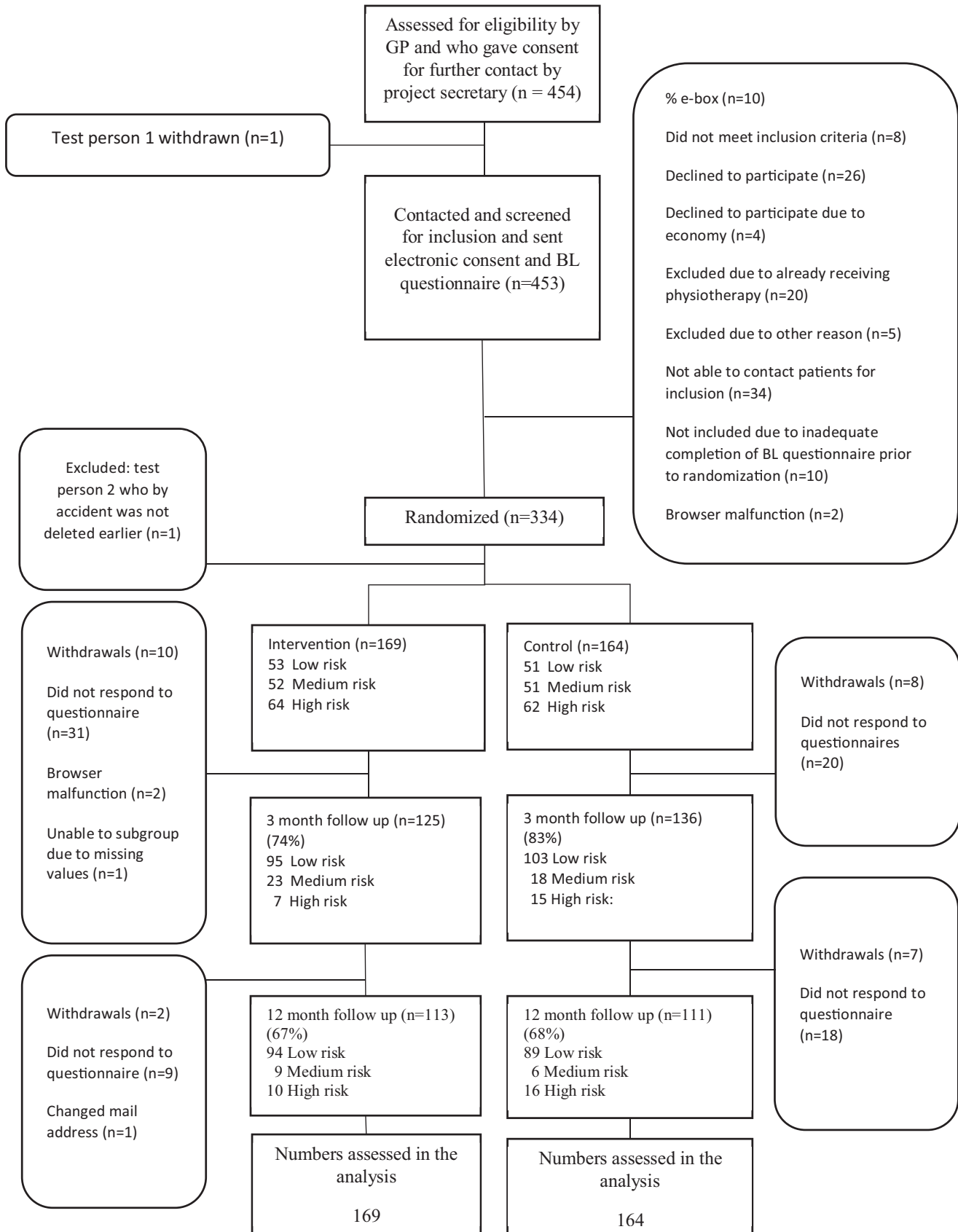
which was developed for use in this study. This was written in Danish and included similar key information to that in *The Back Book* (Burns-Balogh, 2002). Physiotherapists received 5 days' training from a consultant physiotherapist at Keele University to equip them to deliver the appropriate matched treatment. The training was based on the STarT Back trial's high-risk intervention and associated training, and included theory and practical application (Main et al., 2012). In addition, physiotherapists received mentoring sessions in small groups for one year, each session lasted 3 hr and occurred every 6–8 weeks. During the study, the physiotherapists could also use a 'hotline' to telephone experienced clinicians for advice regarding the delivery of the matched treatments. At the initial consultation, the clinicians had access to the completed patient questionnaires including the patient's SBT subgroup allocation. Clinicians were expected to deliver the appropriate matched treatment to each participant (Main et al., 2012). The content of the treatment delivered in the intervention arm is detailed in Appendix 1.

### 2.6 | Current practice (control arm)

For standardization of the care provided in the control arm, the participating physiotherapists and GPs were invited to a 3-hr meeting. During the meeting, the regional LBP guidelines, which included guidance on diagnosis and treatment of LBP and the latest recommendations from international evidence-based guidelines (Danish Society for Internal Medicine, 2006; Regional Management Program, 2010; Savigny et al., 2009) were presented. GPs and physiotherapists in the control practices were asked to provide treatment consistent with these evidence-based guidelines. The clinicians conducting best practice did not receive any specific training on the STarT Back approach and had no access to the study participants' SBT score at any time. Details of the treatment provided to the participants in the intervention and best practice arms can be found in Appendix 1.

### 2.7 | Outcome

Data were collected at baseline, and 3 and 12 months after randomization, and clinician-reported data were collected after initial assessment and after completion of treatment. To have an outcome on the three most common domains of treatment success in LBP (function, generic and objective improvement), the primary outcomes were patient-reported LBP disability measured using the Roland Morris Disability Questionnaire (RMDQ) (Roland & Fairbank, 2000), time off work (days/weeks) and patient-reported global change (single-item rating on 7-point Likert scale). Data on short-term sick leave were obtained via patient self-report. Data



**FIGURE 1** Flow chart of the study

on prolonged sick leave (>2 weeks of consecutive absence) and related social benefits were obtained from the Danish National Register of Public Transfer Payments (DREAM) (Hjollund et al., 2007). Secondary outcomes were pain intensity (11-point numeric rating scale), participant overall satisfaction with improvement (single-item rating), well-being (World Health Organization Well-Being Index [WHO5] [Bech et al., 1996]), participant healthcare resource utilization data from the Danish Nationwide Patient Registry (DNPR) and quality-adjusted life years (QALYs) using the instrument developed by the EurQol group (EQ-5D-5L) (Brooks, 1996; Janssen et al., 2013). For the list of study variables, see Appendix 2.

## 2.8 | Sample size

To mirror the UK study, we used the same RMDQ change score for our sample size calculation as in the UK trial. The calculation showed that 75 patients were needed in each treatment arm based on the subgroup that had the lowest prevalence (high risk) to detect an overall 2.5 between-arm difference in change scores using the RMDQ. A previous study, had suggested that 23% of Danish primary care patients with LBP were at high risk. Therefore, to detect a 2.5 subgroup-level difference, we needed 660 patients in the study.

## 2.9 | Adverse incidents

Clinicians were to register all adverse incidents in the electronic clinician assessment record used in the study.

## 2.10 | Statistical analysis

Results are reported in accordance with CONSORT guidelines (Moher, 2005). Analysis was conducted using intention-to-treat at 3 and 12 months. Between-group primary and secondary continuous outcome scores were compared using appropriate descriptive statistics, multi-level linear models and generalized estimating equation models for categorical outcomes. Analyses were adjusted for cluster effects of treating physiotherapists (Twisk, 2013). For secondary outcome analysis, we used multiple imputations for missing data (Donders et al., 2006; van der Heijden et al., 2006). Sensitivity analysis was based on imputed datasets adjusted for age, sex and clustering by physiotherapists.

The 12-month follow-up was used as the outcome time point for the health economic analysis. Cost data comprised healthcare resource use, and unit costs were obtained from the national provider agreements (agreed prices of healthcare

utilities in primary care) and diagnostic-related groups (agreed prices of healthcare utilities in secondary care) in the DNPR. We extracted data on healthcare utilization from national registries, supplemented with patient-reported data. We calculated QALYs using the EQ-5D-5L with related Danish QALYs weights.

# 3 | RESULTS

## 3.1 | Recruitment

Participants were allocated to either intervention (169) or current practice (164). The study secretary was not able to reach 7% of the patients who were considered eligible by GPs, and 4% who were contacted had already started physiotherapy and were therefore excluded. The mean age of patients excluded from the study was slightly higher than that of patients randomized into the study, and the proportion of females was significantly lower (50.8% vs. 57.7% in the study,  $p < .001$ ). According to the SBT, 31% of study participants were classified as low risk, 31% as medium risk and 38% as high risk. The overall follow-up rate was 79% at 3 months and 67% at 12 months. The follow-up percentage at 3 months differed between groups (intervention 74% vs. current care 83%; see Figure 1).

Although we extended the recruitment period and adjusted the recruitment procedure, GPs only managed to recruit 453 patients to the study. After contact and screening for inclusion, only 334 participants were randomized, which was half the intended number of study participants. This was despite multiple visits by the study team to the GPs in the study, encouragement and reimbursement of €20 per referred patient. To address inadequate recruitment by the GPs, after 18 months we allowed physiotherapists to enrol participants in the study; however, this had little impact on recruitment, as physiotherapists recruited only 15 patients.

## 3.2 | Baseline

Study participants had a median age of 46 years (range 33–57 years), and 58% were women. In terms of highest academic attainment, 28% of study participants were blue-collar workers and 33% had either a bachelor's or master's degree. Almost one-third of participants had experienced LBP for more than 12 months (see Table 1). Patients were moderately disabled with a median score of 12 (8–16) on the RMDQ. They had a median pain severity score of 6 (4–8) on the numeric rating scale (0–10), and one-third used painkillers several times per day. Participants lost to follow-up at 12 months were significantly younger with a mean of 41(30–55) years and scored significantly higher on individual questions of

**TABLE 1** Baseline characteristics of participants. Overall and by risk group

	Overall		Low risk		Medium risk		High risk	
	Total	Intervention	Control	Intervention	Control	Intervention	Control	Intervention
	N = 333	N = 169	N = 164	N = 53	N = 51	N = 52	N = 51	N = 64
Age years	46 (33–57)	46 (35–57)	45 (31–56.5)	46 (31–58)	46 (30–59)	46 (36.5–56)	46 (31–59)	47.5 (38.5–56.5)
Gender, female	192 (57.7%)	103 (60.9%)	89 (54.3%)	30 (56.6%)	27 (52.9%)	30 (57.7%)	31 (60.8%)	43 (67.2%)
Ordinary employed	177 (53.2%)	88 (52.1%)	89 (54.3%)	29 (52.7%)	24 (47.1%)	27 (51.9%)	29 (56.9%)	32 (50%)
Education level								
Primary school	54 (16.4%)	32 (19.0%)	22 (13.6%)	10 (18.9%)	12 (23.5%)	11 (21.2%)	2 (4.1%)	11 (17.5%)
Skilled worker (e.g. blue-collar worker, shop assistant)	92 (27.9%)	42 (25.0%)	50 (30.9%)	10 (18.9%)	10 (19.6%)	15 (28.8%)	22 (44.9%)	17 (27.0%)
Primary school + <3 years	57 (17.3%)	30 (17.9%)	27 (16.7%)	8 (15.1%)	5 (9.8%)	5 (9.6%)	6 (12.2%)	17 (27.0%)
High school +3–4 years	81 (24.5%)	38 (22.6%)	43 (26.5%)	10 (18.9%)	12 (23.5%)	16 (30.8%)	15 (30.6%)	12 (19.0%)
High school >4 years	29 (8.8%)	17 (10.1%)	12 (7.4%)	11 (20.8%)	9 (17.6%)	2 (3.8%)	1 (2.0%)	4 (6.3%)
Other education	17 (5.2%)	9 (5.4%)	8 (4.9%)	4 (7.5%)	3 (5.9%)	3 (5.8%)	3 (6.1%)	2 (3.2%)
Self-reported number of days of work due to back or leg pain in the last 3 months (full time)								
No full days of work	162 (48.6%)	87 (51.5%)	75 (45.7%)	40 (75.5%)	30 (58.8%)	23 (44.2%)	20 (39.2%)	24 (37.5%)
<30 full days of work	105 (31.5%)	47 (27.8%)	58 (35.4%)	4 (7.5%)	12 (23.5%)	18 (34.6%)	19 (37.3%)	25 (39.1%)
>30 full days of work	66 (19.8%)	35 (20.7%)	31 (18.9%)	9 (17.0%)	9 (17.6%)	11 (21.2%)	12 (23.5%)	15 (23.4%)
Self-reported number of days of work due to back or leg pain in the last 3 months (part time)								
No days part time of work	190 (57.1%)	101 (59.8%)	89 (54.3%)	40 (75.5%)	36 (70.6%)	27 (51.9%)	22 (43.1%)	34 (53.1%)
<30 days' part time of work	77 (23.1%)	34 (20.1%)	43 (26.2%)	4 (7.5%)	5 (9.8%)	13 (25.0%)	17 (33.3%)	17 (26.6%)
>30 days' part time of work	66 (19.8%)	34 (20.1%)	32 (19.5%)	9 (17.0%)	10 (19.6%)	12 (23.1%)	12 (23.5%)	13 (20.3%)
Back pain during the last 2 weeks (0–10)	6 (4–8)	6 (4–8)	6 (5–7)	4 (3–6)	4 (3–5)	6 (4–8)	6 (5–7)	8 (7–8)

(Continues)

TABLE 1 (Continued)

	Overall		Low risk		Medium risk		High risk	
	Total	Intervention	Control	Intervention	Control	Intervention	Control	Intervention
Leg pain during the last 2 weeks (0–10)	N = 333	N = 169	N = 164	N = 53	N = 51	N = 52	N = 51	N = 64
	3 (0–6)	3 (0–5)	3 (0–6)	0 (0–2)	1 (0–3)	3 (1–5)	3 (0–6)	5 (2–8)
Duration of current pain:								
<1 month	98 (29.4%)	46 (27.2%)	52 (31.7%)	12 (22.6%)	11 (21.6%)	17 (32.7%)	20 (39.2%)	17 (26.6%)
1–3 months	66 (19.8%)	29 (17.2%)	37 (22.6%)	8 (15.1%)	13 (25.5%)	8 (15.4%)	11 (21.6%)	13 (20.3%)
3–6 months	47 (14.1%)	23 (13.6%)	24 (14.6%)	5 (9.4%)	10 (19.6%)	8 (15.4%)	4 (7.8%)	10 (15.6%)
6–12 months	19 (5.7%)	11 (6.5%)	8 (4.9%)	5 (9.4%)	4 (7.8%)	2 (3.8%)	3 (5.9%)	4 (6.3%)
>12 months	103 (30.9%)	60 (35.5%)	43 (26.2%)	23 (43.4%)	13 (25.5%)	17 (32.7%)	13 (25.5%)	20 (31.3%)
How often did you take pain medication during the last 2 weeks								
Never	56 (16.9%)	33 (19.5%)	23 (14.1%)	20 (37.7%)	18 (35.3%)	6 (11.5%)	4 (8.0%)	7 (10.9%)
Rarely	59 (17.8%)	31 (18.3%)	28 (17.2%)	16 (30.2%)	11 (21.6%)	9 (17.3%)	7 (14.0%)	6 (9.4%)
Once a week	69 (20.8%)	33 (19.5%)	36 (22.1%)	4 (7.5%)	12 (23.5%)	16 (30.8%)	11 (22.0%)	13 (20.3%)
4–6 times a week	19 (5.7%)	9 (5.3%)	10 (6.1%)	2 (3.8%)	0 (0.0%)	3 (5.8%)	4 (8.0%)	4 (6.3%)
Daily	31 (9.3%)	14 (8.3%)	17 (10.4%)	3 (5.7%)	3 (5.9%)	6 (11.5%)	6 (12.0%)	5 (7.8%)
Several times a day	98 (29.5%)	49 (29.0%)	49 (30.1%)	8 (15.1%)	7 (13.7%)	12 (23.1%)	18 (36.0%)	29 (45.3%)
Disturbed sleep during the last 2 weeks due to back- or leg pain (0–10)	5 (3–7)	5 (2–7)	5 (3–7)	3 (1–5)	3 (2–5)	5 (2.5–7)	5 (3–7)	7 (5–8)
RMDQ <sup>a</sup> (0–23)	12 (8–16)	12 (8–16)	12 (8–16)	7 (3–11)	7 (4–9)	12 (10–14)	13 (10–15)	16 (12–19)
Fear Avoidance <sup>b</sup> (0–20)	12 (8–16)	12 (9–15)	12 (8–16)	10 (7–12)	10 (6–13)	12 (9.5–15)	11 (7–15)	14 (10–18)
Controlling and Handling Pain on an average day <sup>c</sup> (0–10)	5 (4–7)	5 (3–7)	5 (4–7)	7 (5–8)	7 (5–8)	5 (5–7)	6 (4–7)	4 (2–5)
When I feel pain, It's terrible and I think it's never going to get any better (0–10) <sup>d</sup>	5 (3–7)	5 (3–7)	5 (3–7)	2 (1–4)	3 (1–5)	5 (3–5)	4 (2–6)	7 (5–8)

(Continues)

TABLE 1 (Continued)

	Overall		Low risk		Medium risk		High risk	
	Intervention	Control	Intervention	Control	Intervention	Control	Intervention	Control
<b>Total</b>	<b>N = 169</b>	<b>N = 164</b>	<b>N = 53</b>	<b>N = 51</b>	<b>N = 52</b>	<b>N = 51</b>	<b>N = 64</b>	<b>N = 62</b>
When I feel pain, I feel I can't stand it anymore (0–10) <sup>d</sup>	3 (1–5)	4 (1–6)	1 (0–2)	1 (0–3)	3 (1–5)	3 (1–6)	5.5 (3.5–7)	6 (4–7)
Risk of the current episode turning into a chronic condition <sup>e</sup>	6 (5–8)	6 (4–8)	5 (2–6)	5 (3–8)	5 (4–8)	5 (3–7)	7 (5–9)	8 (5–9)
WHOS <sup>e</sup> (0–100)	48 (32–68)	48 (36–64)	68 (52–76)	68 (56–72)	52 (36–66)	48 (36–60)	32 (24–48)	36 (24–44)

Note: Data are presented as median (IQR) for continuous measures, and *n* (%) for categorical measures.

<sup>a</sup> Roland Morris Disability Questionnaire.

<sup>b</sup> Sumscore of two questions from the Örebro Musculoskeletal Pain Questionnaire.

<sup>c</sup> Question from the Örebro Musculoskeletal Pain Questionnaire.

<sup>d</sup> Question from Pain Catastrophizing Scale.

<sup>e</sup> WHO5 Questionnaire for Well-being.

catastrophization ( $p = .002$ ), 44% had LBP of >1 year, and the SBT subgroup distribution was low-risk 25%, medium-risk 27% and high-risk 48%.

The odds ratio for poor outcome of 3 months (>7 on RMDQ) in the three risk groups was: low-risk 1.00 (base), medium-risk 1.97 (0.87; 4.46), and high-risk 6.90 (3.01; 15.80), respectively. The proportion of patients having disabilities above 7 on the RMDQ at 3 months was 26.9% in the low-risk groups compared to 74.6% in the high-risk groups.

There was a slightly higher proportion of women within the two trial arms (intervention 60.9% vs. control 54.3%). There were no significant differences between the two arms in any other baseline parameters.

### 3.3 | Primary outcomes

Participants in both the intervention and control arms were improved in the primary outcome parameters at 3 months. The mean change in the primary outcome of disability was similar in the intervention and control arms at 3 months (5.9 [*SD* 5.6] vs. 5.5 [5.8]), resulting in a non-significant mean difference between arms of  $-0.5$  (95% CI  $-1.8$  to  $0.9$ ), and at 12 months (6.1 [6.1] vs. 6.5 [5.8]), with a mean difference of  $0.1$  (95% CI  $-1.5$  to  $1.6$ ). Differences in outcomes between the intervention and control arm by subgroup were similar to the overall results (see Table 2).

There was no overall difference between arms in self-reported full-time or part-time sick leave, due to back pain at 3 months. In contrast, self-reported full-time sick leave was lower in the low-risk subgroup of the intervention arm ( $-3.75$  [95% CI  $-7.49$  to  $-0.02$ ]), but at 12 months, these differences were no longer present. We found no differences in the medium- and high-risk groups. Analysis of data from the DREAM on long-term sick leave showed no differences in number of weeks of sick leave with compensation across the two study arms (intervention 3.6 weeks vs. control group 4.0 weeks during the 12-month follow-up period). The work participation score (WPS), defined as a number of return to work (RTW) weeks divided by RTW weeks + number of weeks receiving social transfer payments (Biering et al., 2013), was 89% in the intervention arm and 87.5% in the control arm (see Appendix 3).

The final primary outcome of patient-reported global change showed overall small differences across arms. At 3 months, 7% were worse, 39% had no change and 55% were better or much better in the intervention arm, and 5% were worse, 35% had no change and 60% were better or much better in the control arm. At 12 months, differences between the intervention and the control arms were as follows: worse 6% versus 9%, no change 32% versus 32% and better/much better 61% versus 59% (see Table 2).



**TABLE 2** Mean change, differences and proportions on primary outcomes at 3 and 12 months

	All participants			Low-risk participants			Medium-risk participants			High-risk participants		
	Intervention group	Control group	<i>p</i> -value	95% CI	<i>p</i> -value	95% CI	95% CI	<i>p</i> -value	95% CI	95% CI	<i>p</i> -value	
	169	164		53	51	52	64	62				
3 months												
RMDQ												
Intervention group (SD)	5.93	-5.60		4.45	5.13	7.00	5.42	6.73	6.00			
Control group (SD)	5.52	5.89		4.16	3.74	5.91	5.72	6.45	7.46			
Mean difference	-0.45	-1.82	.52	-0.47	-2.53	1.58	1.86	0.36	-2.63	3.35	.81	
Sensitivity analysis <sup>a</sup>	-0.40	-2.10	.63	-0.53	-2.75	1.68	2.39	0.33	-2.47	3.14	.81	
Self-reported full time off work												
Intervention group (SD)	-1.48	13.70		0.26	1.54	-1.15	9.36	-3.78	22.07			
Control group (SD)	-1.88	13.58		-3.09	13.13	0.20	14.71	-2.77	12.88			
Mean difference	-0.49	-4.26	.80	-3.75	-7.49	-0.02	9.73	0.21	-11.09	11.51	.97	
Sensitivity analysis <sup>a</sup>	0.10	-3.70	.96	-2.99	-7.42	1.45	9.90	0.50	-8.62	9.61	.91	
Self-reported part time off work												
Intervention group (SD)	-1.12	12.39		0.11	0.92	-1.74	15.50	-2.00	16.17			
Control group (SD)	0.18	13.24		-0.33	10.20	1.84	13.78	-0.98	15.30			
Mean difference	1.76	-2.24	.38	-0.32	-3.60	2.96	13.29	3.38	-6.51	13.27	.49	
Sensitivity analysis <sup>a</sup>	1.80	-1.98	.34	-0.49	-3.90	3.81	11.95	2.00	-6.52	10.53	.63	
Self-reported global change <sup>b</sup> , <i>n</i> (%)												
Worse	8 (6.6)	6 (4.6)		2 (4.3)	0 (0.0)	2 (5.7)	3 (6.8)	4 (9.7)	3 (6.8)			
No change	47 (38.5)	46 (35.1)		17 (36.9)	16 (37.2)	13 (37.1)	11 (25.0)	17 (41.5)	19 (43.2)			
Better	26 (21.3)	34 (26.0)		9 (19.6)	13 (30.2)	11 (31.4)	12 (27.3)	6 (14.6)	9 (20.5)			
Much better	41 (33.6)	45 (34.4)		18 (39.1)	14 (32.6)	9 (25.7)	18 (40.9)	14 (34.1)	13 (29.5)			
12 months												
RMDQ												
Intervention group (SD)	6.09	6.10		4.13	4.84	8.16	5.66	6.42	7.06			

(Continues)

TABLE 2 (Continued)

	All participants		Low-risk participants		Medium-risk participants		High-risk participants	
	Intervention group	Control group	Intervention group	Control group	Intervention group	Control group	Intervention group	Control group
Control group (SD)	6.50	5.83	4.26	4.21	6.22	5.41	9.23	6.73
Mean difference	0.06	-1.49	-0.34	-2.57	-2.19	-4.58	3.13	-0.11
Sensitivity analysis <sup>a</sup>	0.07	-1.48	-0.77	-3.19	-0.63	-3.57	1.91	-0.89
Self-reported full time off work								
Intervention group (SD)	-0.21	14.25	-1.95	14.45	-1.94	12.05	3.08	15.59
Control group (SD)	1.53	9.54	1.18	2.76	1.17	3.09	2.25	16.28
Mean difference	0.94	-1.45	0.77	-0.33	3.28	-1.04	-2.95	-9.11
Sensitivity analysis <sup>a</sup>	1.32	-1.21	-0.25	-3.46	2.33	-3.35	-1.11	-6.52
Self-reported part time off work								
Intervention group (SD)	-0.69	11.62	-2.11	14.34	-0.44	11.57	0.50	8.35
Control group (SD)	1.45	9.64	1.21	5.40	-0.89	8.31	4.06	13.31
Mean difference	1.65	-0.79	1.06	-1.02	-0.12	-5.26	3.54	-2.42
Sensitivity analysis <sup>a</sup>	1.48	-0.94	-0.48	-3.18	-0.42	-5.80	2.80	-2.33
Self-reported global change <sup>b</sup> , n (%)								
Worse	7 (6.3)	10 (9.1)	1 (2.6)	2 (5.3)	4 (11.8)	7 (19.4)	2 (5.3)	1 (2.8)
No change	36 (32.4)	35 (31.8)	11 (28.2)	12 (31.6)	10 (29.4)	10 (29.4)	15 (39.5)	13 (36.1)
Better	30 (27.0)	27 (24.6)	12 (30.8)	10 (26.3)	8 (23.5)	7 (19.4)	10 (26.3)	10 (29.4)
Much better	38 (34.2)	38 (34.6)	15 (38.5)	14 (36.8)	12 (35.3)	12 (33.3)	11 (28.9)	12 (33.3)

<sup>a</sup> Based on imputed datasets adjusted for age, sex, clustering by physiotherapist.<sup>b</sup> The 7-point Likert scale was collapsed into four categories.

### 3.4 | Secondary outcomes

The analyses of secondary outcomes are presented in Table 3. The overall between-group mean differences for back pain severity and leg pain severity were 0.16 (95% CI  $-0.46$  to  $0.78$ ) and  $0.34$  ( $-0.52$  to  $1.17$ ) at 3 months and  $-0.45$  ( $-1.24$  to  $0.34$ ) and  $-0.16$  ( $-1.6$  to  $0.74$ ) at 12 months, respectively.

At 3-months' follow-up, participants rated their overall satisfaction with their symptoms by using the single item 'Would you be satisfied if your condition remained as it is at this time point'? Overall, 36% were satisfied in the intervention group and 33% in the control group ( $p = .18$ ); at 12 months, the percentages were 34% and 32%, respectively.

Patient-reported well-being on the WHO-5 showed an overall improvement of  $\pm 15$  points in both arms, with a mean group difference of  $-1.3$  (95% CI  $-5$  to  $4.0$ ) at 3 months, and a 16.2- versus 14.7-point improvement in the intervention and control arms, respectively, with a mean group difference of  $1.9$  ( $-3.5$  to  $7.3$ ) at 12 months.

The mean number of physiotherapist-reported treatment sessions received by study participants differed between the two treatment arms, with intervention arm patients receiving fewer appointments, with a median of two (interquartile range [IQR] 2–5) compared with the control arm, with a median of four (IQR 2–6,  $p = .002$ ). This overall between-arm difference was driven in part by the medium-risk subgroup treatment sessions, with a median of three in the intervention arm (IQR 2–6) and four (IQR 3–8) in the control arm ( $p = .003$ ). However, the difference was greatest between the low-risk subgroups, with low-risk patients receiving a median of two (IQR 1–2) treatment sessions in the intervention arm compared with three (IQR 2–4) in the control arm ( $p < .001$ ) (See Appendix 4).

Healthcare utilization in primary care was very similar across the two study arms. In Table 4, the mean number of GP consultations per participant is displayed:  $8.7$  ( $SD$  9.2) and  $7.6$  ( $6.2$   $SD$ ) for the intervention and control arms, respectively. Also displayed is the chiropractor appointment rate:  $4.7$  ( $SD$  3.8) and  $5.9$  ( $SD$  4.7). Nine patients from the intervention arm and 18 from the control arm were referred to a specialized spine centre. There were no differences between the two arms in referrals to imaging.

Healthcare expenses for utilities in the trial reflect the mean regional healthcare expenses in euro for each treatment study arm (e.g. the mean regional expense for one physiotherapy treatment in the intervention arm is €13.08, which is €1.19 less than the mean healthcare expense in the control arm, €14.26). In terms of physiotherapy services, GP services and prescribed opioids, there were significantly lower costs in the intervention arm. The difference in total healthcare costs during the 12 months prior to inclusion compared with 12 months after inclusion showed an increased cost of €13.36 ( $SD$  528.59) in the intervention arm and €228.51 ( $SD$

829.8) in the control arm (see Appendix 5). These differences seem to be primarily driven by less prescription of pain medication by GPs and lower x-ray expenses in the intervention group (see Table 4).

QALYs estimates based on the EQ-5D-5L resulted in small additional health benefits in both arms, with a gain of  $0.12$  (95% CI  $0.08$ – $0.16$ ) QALYs in the intervention group and  $0.14$  ( $0.11$ – $0.18$ ) in the control arm, but this was not significant between arms ( $p = .31$ ; see Appendix 6).

### 3.5 | Harm

Overall, 14 incidents were reported, five in the intervention arm and nine in the control arm. All incidents were described as minor and were categorized as 'exercise soreness or flair up of known pain'.

## 4 | DISCUSSION

The trial found no statistical differences in primary or secondary outcomes between stratified and usual care. Stratified care reduced the number of clinician-reported treatment sessions and resulted in lower healthcare costs, but did not result in greater benefits in terms of cost-effectiveness.

Medium-risk participant satisfaction was lower in the control than the intervention arm at 3 months. This was surprising as the items in the SBT that characterize the medium-risk subgroup is mostly physical, and the medium-risk-matched treatment broadly resembles current care. We expected to see differences in high-risk participant satisfaction, providing different treatment for the intervention group, or in low-risk intervention patients, who had fewer treatment sessions. Subgroup findings must be interpreted cautiously, however, due to the small sample size.

An overall significant difference existed in the number of clinician-reported treatment sessions across the 2 study arms, with participants in the intervention arm receiving fewer treatments. This finding is in contrast to the findings by Beneciuk and George (Beneciuk & George, 2015). Their study showed no differences in the numbers of healthcare visits for the low- and medium-risk subgroups. However, high-risk participants who received stratified care had more visits compared to those receiving standard care. This is in line with the STarT Back Trial results. The differences in our study were mainly driven by fewer sessions in the low-risk group. We were not able to identify the duration of treatment sessions, but the reduced physiotherapy costs could indicate a shorter duration as well as fewer sessions. Using the SBT risk profile to guide decision making might have provided clinicians with confidence to discharge low-risk patients sooner, thereby avoiding overtreatment. The



TABLE 3 (Continued)

	All participants			Low-risk participants			Medium-risk participants			High-risk participants		
		95% CI	<i>p</i> -value		95% CI	<i>p</i> -value		95% CI	<i>p</i> -value		95% CI	<i>p</i> -value
<b>Intervention group</b>	<b>169</b>			<b>53</b>			<b>52</b>			<b>64</b>		
<b>Control group</b>	<b>164</b>			<b>51</b>			<b>51</b>			<b>62</b>		
12 months												
Back Pain												
Intervention group ( <i>SD</i> )	2.75	2.61		2.08	2.27		2.94	2.55		3.26	2.88	
Control group ( <i>SD</i> )	2.51	2.96		2.11	2.46		1.82	2.99		3.67	3.14	
Mean difference	-0.45	-1.24	.26	-0.17	-1.32	.98	-1.21	-3.12	.70	0.13	-1.26	1.51
Leg Pain												
Intervention group ( <i>SD</i> )	1.40	2.96		0.15	1.76		1.81	3.08		2.28	3.42	
Control group ( <i>SD</i> )	1.26	2.85		0.45	1.93		0.87	3.31		2.53	2.78	
Mean difference	-0.16	-1.06	.73	0.16	-0.88	1.20	-1.03	-2.56	0.51	0.45	-1.14	2.04
Disturbed Sleep												
Intervention group ( <i>SD</i> )	2.22	3.17		1.38	2.55		2.35	3.80		2.95	3.01	
Control group ( <i>SD</i> )	2.47	3.65		2.18	2.81		1.94	4.60		3.31	3.32	
Mean difference	0.16	-0.88	1.21	0.65	-0.72	2.01	-0.99	-3.27	1.29	0.71	-0.84	2.25
WHO5 Well-being												
Intervention group ( <i>SD</i> )	-16.18	18.05		-11.47	15.76		-15.52	15.26		-21.47	21.20	
Control group ( <i>SD</i> )	-14.72	22.49		-7.16	16.85		-10.56	23.01		-27.20	22.58	
Mean difference	1.93	-3.47	7.34	4.97	-0.68	10.62	4.98	-4.59	14.55	-4.37	-15.79	7.05

**TABLE 4** Overview of primary care, secondary care and prescribed medication healthcare utilization and associated costs

	Resources			Costs <sup>a</sup>			<i>p</i> -value
	Intervention		Control	Intervention		Control	
	Number	Mean (SD)	Number	Mean (SD)	Mean (SD)		
<b>Primary Care</b>							
Physiotherapy	125	7.1 (8.5)	141	6.4 (6.8)	13.08 [12.76–13.41]	14.26 [13.91–14.61]	.000
General practitioner	159	8.7 (9.2)	148	7.6 (6.2)	12.67 [12.31–13.03]	13.91 [13.47–14.35]	.000
Chiropractor	18	4.7 (3.8)	16	5.9 (4.7)	7.02 [6.31–7.74]	6.38 [5.77–6.99]	.178
Rehabilitation (municipality)	9	17.6 (16.2)	10	19.1 (15.5)	16.00 [16.00–16.00]	16.00 [16.00–16.00]	NA
Other primary care contacts <sup>b</sup>	11	2.8 (1.3)	8	1.9 (1.1)	41.76 [31.95–51.57]	50.20 [33.24–67.15]	.381
<b>Secondary Care</b>							
Spine centre	9	2.9 (2.2)	18	3.0 (2.8)	489.87 [–81.19–1060.93]	470.03 [89.83–850.24]	.954
MRI	31	1.2 (0.4)	29	1.1 (0.4)	365.29 [290.73–439.85]	328.55 [257.54–399.56]	.485
CT imaging	8	1.3 (0.7)	6	1.3 (0.8)	439.12 [74.86–803.38]	490.52 [35.60–945.43]	.862
X-ray	55	1.6 (1.1)	53	1.5 (0.8)	310.06 [46.36–573.77]	981.42 [90.51–1872.33]	.069
Other secondary care contacts <sup>c</sup>	70	4.9 (5.7)	67	4.7 (5.4)	440.91 [335.13–546.70]	455.25 [341.07–569.42]	.857
<b>Medication</b>							
Prescribed non-opioid analgesics	76	3.59 (3.55)	72	3.78 (4.70)	7.84 [7.24–8.45]	7.03 [6.49–7.58]	.051
Prescribed opioid analgesics	34	5.65 (10.45)	27	4.96 (5.40)	7.71 [6.25–9.17]	16.61 [12.85–20.37]	.000
Prescribed tricyclic medication <sup>d</sup>	21	5.10 (5.01)	27	7.26 (6.63)	22.45 [14.96–29.94]	30.85 [23.25–38.46]	.133

*Note:* The table shows mean number of treatments (and standard deviation) for participants in each treatment group. For all participants who has a registered treatment in the Danish Nationwide Patient Registry.

Costs reflect the mean regional healthcare expenditure in EUR for each treatment given in the two treatments groups.

<sup>a</sup> €.

<sup>b</sup> Rheumatologist, Neurologist.

<sup>c</sup> e.g. emergency care, surgery (only one patient from the control arm ended up having back surgery).

<sup>d</sup> used for neurogenic pain.

subgroup classification of the SBT reflects items clinicians usually use to assess LBP (Hill et al., 2008), and although Kent et al., (2009) found wide variability in the assessment of non-specific LBP (Kent et al., 2009), clinicians primarily use physical examination findings to support clinical decision making. Although other decision tools are well known in physiotherapy to support treatment plans or determine severity (Brazier et al., 1992; Foster et al., 2013; Freynhagen et al., 2006), recent studies and guidelines indicate that multiple factors play an important predictive role (Beneciuk et al., 2013; Stochkendahl et al., 2018; Trinderup et al., 2018). The findings of this study indicate that SBT subgroup classification seems useful and can support clinicians in reducing the number of treatment sessions for some patients, without affecting overall satisfaction with the quality of care.

We used 2 different measures to monitor the amount of sick leave. Although no overall differences between the 2 arms were detected, the 2 measures displayed different patterns. A large proportion of study participants reported short-term sick leave at 3 and 12 months, but this could not be tracked in the register data. The DREAM only registers social transfer payments after 2 weeks of continuous sick leave. According to the DREAM data, patients had prolonged periods of work without shifting between receiving social transfer payments and work participation in both study arms; this pattern is not typical for LBP patients (Hestbaek et al., 2003; Kongsted et al., 2016). Studies investigating trajectories show short periods of days off work as indicated by self-reported data (Hestbaek et al., 2003; Kongsted et al., 2016). This indicates that using DREAM data to measure sick leave for this population might not be optimal. The WPS was equally high in both arms ( $\pm 88\%$ ). This indicates that most participants remained at work despite LBP. This is in line with earlier studies (Chen et al., 2018; Forsbrand et al., 2018; Grovle et al., 2013). It could be argued that RTW might not be the right outcome measure for intervention studies where participants are working to some extent and that short-term sick leave, for example, measured in days is more likely to detect change or between group differences, than measured in weeks.

The lower healthcare costs at 12 months in the intervention group were due to lower utilization of physiotherapy services, GP services, pain medication and to some extent x-ray expenses. We were not able to determine why healthcare expenses were lower for physiotherapists and GPs, besides the above speculation on shorter durations of treatment sessions. The increased cost of x-rays in the control arm might have been because X-rays are performed during hospital admission (e.g. acute admission due to progressing disc herniation), however, this is speculation we were not able to disentangle healthcare expenses in these cases.

## 4.1 | Strengths and limitations

Our findings did not reflect the results of the STarT Back trial. Our findings did not reflect the results of the STarT Back trial. The findings were more in line with the US MATCH study (Cherkin et al., 2018) and the TARGET study (Delitto et al., 2021), in which the intervention arm did not have a significant effect on patient outcomes, healthcare use or the development of chronic LBP, compared to the control arm.

The findings were more in line with the US MATCH study (Cherkin et al., 2018), in which the intervention arm did not have a significant effect on patient outcomes or healthcare use, compared to the control arm. As we more closely tried to replicate the original STarT Back trial, therefore, the explanations of the results are likely to be different.

We aimed to include participants 10–14 days after they consulted with their GP. However, we found that project physiotherapists assessed some participants as early as 3–5 days after consulting their GP (typically 24–48 hr after onset of the LBP episode). A previous study suggested that assessing participants this early might affect the predictive ability of the SBT (Field & Newell, 2012). The predictive ability in this study was slightly lower compared to an earlier study in Danish primary care (Morso et al., 2013) and it could potentially affect the results, but still, the SBT subgroup classification was predictive of prolonged disability. A recent study shows that stratification at 6 weeks is an optimal time point (Medeiros et al., 2018); this is in line with earlier studies by Khan et al. (Khan et al., 2019) and Morso et al. (Morso et al., 2016). LBP duration might have affected the SBT predictive ability, perhaps risking subgroup misclassification. The SBT subgroup distribution for participants lost to follow-up at 12 months was different from those with complete data. Compared to participants with complete data, proportionally more of those lost to follow-up were high-risk, they also tended to be younger and had a higher level of catastrophizing. Subjects lost to follow-up were evenly distributed across the 2 treatment arms. Therefore, these differences are not likely to have affected the results.

The physiotherapists in the intervention arm were experienced clinicians (mean experience >7 years) and received additional training in delivering the matched treatments. Due to slow recruitment, however, these clinicians reported insufficient opportunities to practise and develop their confidence and competence in delivering the appropriate matched treatments. We held ongoing workshops during the study inclusion period to support clinicians. However, this may have been inadequate. In addition, we extended the recruitment period because of recruitment difficulties, and physiotherapists reported weariness in relation to the prolonged duration of the study. Although we paid frequent visits to the participants, motivation might have deteriorated over time, affecting fidelity to the matched treatments. In contrast, we

aimed to standardize treatment in the control arm by offering 3-hr training to GPs and physiotherapists. This might have increased the quality of current care provided to participants in the control arm with a small wash-out effect as result.

Although fewer physiotherapy sessions were delivered to low-risk participants in the intervention arm, the content of the treatment sessions for participants in the low-risk subgroup in the 2 arms is likely to have been similar. In addition, the matched treatment for participants in the medium subgroup usually resembles best practice for routine physiotherapy care. Therefore, the biggest difference in the treatments provided between the 2 arms was expected for the high-risk subgroup. Though, physiotherapists providing the control arm were blinded to participants' SBT scores and were asked not to deliver stratified care, all physiotherapists are likely to have had some knowledge of the SBT approach, prior to the study. The development work of the Danish SBT version (Morso et al., 2011; Morso et al., 2013) and the publication of the STarT Back trial (Hill et al., 2011) had promoted the approach in Danish physiotherapy circles. Furthermore, in the period from publication of the SBT to the conduction of this study, increased focus on psychosocial factors might have changed clinical practice. This might have washed out some of the potential effectiveness of the matched treatments, particularly in the high-risk subgroup.

In contrast to the STarT Back trial, where clinicians delivered the different matched treatments, the same physiotherapists in this study delivered all three (low-, medium- and high-risk) matched treatments. This was done to better reflect clinical practice in Denmark but could have increased the risk of matched treatments not being delivered as intended, especially for the matched high-risk treatment. Therefore, it could be argued that some of the clinicians in the intervention arm may not have delivered the appropriate matched treatment. Despite a systematic effort in our study to ensure fidelity, we found that the extent to which stratified care was provided by clinicians varied. This is in keeping with a study conducted in the United States which outlined the difficulties of implementing stratified care into the US healthcare system (Hsu et al., 2019). Although this is contradicted by the fewer sessions in the low-risk group, it could be a potential limitation of this study.

In the study, we failed to recruit a sufficient number of participants via GPs despite payment per recruited patient and several visits from the primary investigator. We randomized at the patient level, but had we randomized at the practice level, we may have recruited more participants. Furthermore, the response rate at 12 months was only moderate, but the outcome was imputed by multiple imputation. In the study, we found a small between arm mean differences in the primary outcome (RMDQ) of  $-0.5$ . Although only 334 of the intended participants were recruited, the overall powering of the study was sufficient to detect between arm differences.

The primary outcome was measured with considerable precision (95% CI  $-1.8$  to  $0.9$ ) with the lower limit of the confidence interval well above the minimal important difference of  $-2.5$  points. Therefore, the low inclusion did not affect the confidence of the overall results in the study. Still, this study allows us only to conclude on the overall between-group results, whereas subgroups analyses should be interpreted with caution.

## 4.2 | Generalizability

The total number of patients recruited was much lower than expected. However, we believe participants to be generalizable to patients with non-specific LBP in Danish primary care, as we do not believe there has been any systematic selection bias (Morso et al., 2013). As patients lost to follow-up were evenly distributed in the arms, we assume the risk of attrition bias to be limited. The fact, that patients lost to follow-up were more frequently at higher risk of persistent disability and had higher scores for pain catastrophizing compared to the analysed patient sample, might reduce the generalizability of the results.

The clinician training delivered as part of this trial can be generalized to other settings. Although stratified care may be challenging to deliver, the SBT appears to have usefully informed clinical decision making; this is in keeping with a previous study (Brunner et al., 2018). In future, we would recommend assessing fidelity to the matched treatments and monitoring the number of treatment sessions provided.

## 5 | CONCLUSION

There were no differences in clinical outcomes between patients with non-specific LBP, who received stratified care, compared to usual care in Danish primary care. However, stratified care may be associated with fewer treatment sessions and reduced healthcare costs related to the prescription of pain medication and x-rays.

## ACKNOWLEDGEMENTS

The authors thank all GPs, physiotherapists and patients who participated in this trial. The authors also acknowledge the work done to enrol patients by the project secretaries and clinic secretaries. We acknowledge Søren Bie Bogh, OPEN; Nana Hyldig, OPEN; and Marie Kruuse, DaCHE for the valuable help in data analysis and thank Jonathan Hill for valuable input though the trial.

## CONFLICTS OF INTEREST

All authors declare no conflict of interest in the study.








## AUTHORS' CONTRIBUTIONS

LM and DHC contributed to the conception, overall design and funding of the study. Both were actively involved in managing the trial. BSC, KOR and JS contributed to the study conception and design. LM wrote the first draft of the manuscript. All authors participated in the critical revision of this manuscript and approved the final edition. All authors gave consent for publication.

## ETHICS AND REGISTRATION

Participation was based on informed consent. All patients in the study received written and verbal information regarding the study. Patients who withdrew consent or did not want to participate in the project were referred and treated in concordance with usual practice in primary care. The study was approved by the Scientific Ethics Committee of Southern Denmark (file no. S-20140205) and the Danish Data Protection Agency (file no. 15/3321). All rules of storage of personal information were met according to the Danish Data Protection Agency. Data management was conducted according to the regional rules of the Region of Southern Denmark. Trial registration: ClinicalTrials.gov/NCT02612467 (Registered 16 November 2015).

## ORCID

Lars Morsø  <https://orcid.org/0000-0003-3084-4186>  
 Kim Olsen Rose  <https://orcid.org/0000-0002-0473-4736>  
 Berit Schiøttz-Christensen  <https://orcid.org/0000-0003-3866-186X>  
 Jens Søndergaard  <https://orcid.org/0000-0002-1629-1864>  
 David H. Christiansen  <https://orcid.org/0000-0001-7458-3921>

## REFERENCES

- Bech, P., Gudex, C., & Johansen, K. S. (1996). The WHO (ten) well-being index: Validation in diabetes. *Psychotherapy and Psychosomatics*, *65*, 183–190. <https://doi.org/10.1159/000289073>
- Beneciuk, J. M., Bishop, M. D., Fritz, J. M., Robinson, M. E., Asal, N. R., Nisenzon, A. N., & George, S. Z. (2013). The STarT back screening tool and individual psychological measures: Evaluation of prognostic capabilities for low back pain clinical outcomes in outpatient physical therapy settings. *Physical Therapy*, *93*, 321–333. <https://doi.org/10.2522/ptj.20120207>
- Beneciuk, J. M., & George, S. Z. (2015). Pragmatic implementation of a stratified primary care model for low back pain management in outpatient physical therapy settings: Two-phase, sequential preliminary study. *Physical Therapy*, *95*, 1120–1134.
- Biering, K., Hjøllund, N. H., & Lund, T. (2013). Methods in measuring return to work: A comparison of measures of return to work following treatment of coronary heart disease. *Journal of Occupational Rehabilitation*, *23*, 400–405. <https://doi.org/10.1007/s10926-012-9405-x>
- Brazier, J. E., Harper, R., Jones, N. M., O' Cathain, A., Thomas, K. J., Usherwood, T., & Westlake, L. (1992). Validating the SF-36 health survey questionnaire: New outcome measure for primary care. *BMJ*, *305*(6846), 160–164. <https://doi.org/10.1136/bmj.305.6846.160>
- Brooks, R. (1996). EuroQol: The current state of play. *Health Policy*, *37*(1), 53–72. [https://doi.org/10.1016/0168-8510\(96\)00822-6](https://doi.org/10.1016/0168-8510(96)00822-6)
- Brunner, E., Dankaerts, W., Meichtry, A., O'Sullivan, K., & Probst, M. (2018). Physical therapists' ability to identify psychological factors and their self-reported competence to manage chronic low back pain. *Physical Therapy*, *98*, 471–479. <https://doi.org/10.1093/ptj/pzy012>
- Burns-Balogh, P. (2002). *The back book*. (2nd ed.). TSO.
- Chen, Y., Campbell, P., Strauss, V. Y., Foster, N. E., Jordan, K. P., & Dunn, K. M. (2018). Trajectories and predictors of the long-term course of low back pain: Cohort study with 5-year follow-up. *Pain*, *159*, 252–260. <https://doi.org/10.1097/j.pain.0000000000001097>
- Cherkin, D., Balderson, B., Brewer, G., Cook, A., Estlin, K. T., Evers, S. C., Foster, N. E., Hill, J. C., Hawkes, R., Hsu, C., Jensen, M., LaPorte, A.-M., Levine, M. D., Piekara, D., Rock, P., Sherman, K., Sowden, G., Wellman, R., & Yeoman, J. (2016). Evaluation of a risk-stratification strategy to improve primary care for low back pain: The MATCH cluster randomized trial protocol. *BMC Musculoskeletal Disorders*, *17*, 361. <https://doi.org/10.1186/s12891-016-1219-0>
- Cherkin, D., Balderson, B., Wellman, R., Hsu, C., Sherman, K. J., Evers, S. C., Hawkes, R., Cook, A., Levine, M. D., Piekara, D., Rock, P., Estlin, K. T., Brewer, G., Jensen, M., LaPorte, A.-M., Yeoman, J., Sowden, G., Hill, J. C., & Foster, N. E. (2018). Effect of low back pain risk-stratification strategy on patient outcomes and care processes: The MATCH randomized trial in primary care. *Journal of General Internal Medicine*, *33*, 1324–1336. <https://doi.org/10.1007/s11606-018-4468-9>
- Danish Society for Internal Medicine (DSAM). (2006) Diagnostics and treatment in general practice organisation for general practitioners, 2006. National Guideline.
- Delitto, A., Patterson, C. G., Stevans, J. M., Freburger, J. K., Khoja, S. S., Schneider, M. J., Greco, C. M., Freel, J. A., Sowa, G. A., Wasan, A. D., Brennan, G. P., Hunter, S. J., Minick, K. I., Wegener, S. T., Ephraim, P. L., Beneciuk, J. M., George, S. Z., & Saper, R. B. (2021). Stratified care to prevent chronic low back pain in high-risk patients: The TARGET trial. A multi-site pragmatic cluster randomized trial. *EclinicalMedicine*, *34*, 100795. <https://doi.org/10.1016/j.eclinm.2021.100795>
- Donders, A. R., van der Heijden, G. J., Stijnen, T., & Moons, K. G. (2006). Review: A gentle introduction to imputation of missing values. *Journal of Clinical Epidemiology*, *59*, 1087–1091. <https://doi.org/10.1016/j.jclinepi.2006.01.014>
- Field, J., & Newell, D. (2012). Relationship between STarT back screening tool and prognosis for low back pain patients receiving spinal manipulative therapy. *Chiropractic & Manual Therapies*, *20*, 17. <https://doi.org/10.1186/2045-709X-20-17>
- Flynn, T., Fritz, J., Whitman, J., Wainner, R., Magel, J., Rendeiro, D., Butler, B., Garber, M., & Allison, S. (2002). A clinical prediction rule for classifying patients with low back pain who demonstrate short-term improvement with spinal manipulation. *Spine*, *27*(24), 2835–2843. <https://doi.org/10.1097/00007632-200212150-00021>
- Forsbrand, M. H., Grahn, B., Hill, J. C., Petersson, I. F., Post Sennehed, C., & Stigmar, K. (2018). Can the STarT back tool predict health-related quality of life and work ability after an acute/subacute episode with back or neck pain? A psychometric validation study in primary care. *British Medical Journal Open*, *8*, e021748. <https://doi.org/10.1136/bmjopen-2018-021748>
- Foster, N. E., Hill, J. C., & Hay, E. M. (2011). Subgrouping patients with low back pain in primary care: Are we getting any better at it? *Manual Therapy*, *16*, 3–8. <https://doi.org/10.1016/j.math.2010.05.013>

- Foster, N. E., Hill, J. C., O'Sullivan, P., & Hancock, M. (2013). Stratified models of care. *Best Practice & Research: Clinical Rheumatology*, 27, 649–661. <https://doi.org/10.1016/j.berh.2013.10.005>
- Foster, N. E., Mullis, R., Hill, J. C., Lewis, M., Whitehurst, D. G. T., Doyle, C., Konstantinou, K., Main, C., Somerville, S., Sowden, G., Wathall, S., Young, J., & Hay, E. M. (2014). Effect of stratified care for low back pain in family practice (IMPACT Back): A prospective population-based sequential comparison. *Annals of Family Medicine*, 12, 102–111. <https://doi.org/10.1370/afm.1625>
- Freyenhagen, R., Baron, R., Gockel, U., & Tolle, T. R. (2006). pain-DETECT: A new screening questionnaire to identify neuropathic components in patients with back pain. *Current Medical Research and Opinion*, 22, 1911–1920. <https://doi.org/10.1185/030079906X132488>
- Fritz, J. M., Cleland, J. A., & Childs, J. D. (2007). Subgrouping patients with low back pain: Evolution of a classification approach to physical therapy. *The Journal of Orthopaedic and Sports Physical Therapy*, 37, 290–302. <https://doi.org/10.2519/jospt.2007.2498>
- Fritz, J. M., & George, S. (2000). The use of a classification approach to identify subgroups of patients with acute low back pain. Interrater reliability and short-term treatment outcomes. *Spine*, 25(1), 106–114. <https://doi.org/10.1097/00007632-200001010-00018>
- Groble, L., Haugen, A. J., Keller, A., Ntvig, B., Brox, J. I., & Grotle, M. (2013). Prognostic factors for return to work in patients with sciatica. *The Spine Journal*, 13, 1849–1857. <https://doi.org/10.1016/j.spinee.2013.07.433>
- Hartvigsen, L., Kongsted, A., Vach, W., Salmi, L. R., & Hestbaek, L. (2018). Does a diagnostic classification algorithm help to predict the course of low back pain? A study of Danish chiropractic patients with 1-year follow-up. *The Journal of Orthopaedic and Sports Physical Therapy*, 48, 837–846. <https://doi.org/10.2519/jospt.2018.8083>
- Hestbaek, L., Leboeuf-Yde, C., & Manniche, C. (2003). Low back pain: What is the long-term course? A review of studies of general patient populations. *European Spine Journal*, 12, 149–165. <https://doi.org/10.1007/s00586-002-0508-5>
- Hill, J. (2008). *Identifying subgroups among patients with low back pain in primary care: Evaluating the STarT Back Tool (Unpublished thesis)*. Primary Care and Health Sciences, Keele University.
- Hill, J. C., Dunn, K. M., Lewis, M., Mullis, R., Main, C. J., Foster, N. E., & Hay, E. M. (2008). A primary care back pain screening tool: Identifying patient subgroups for initial treatment. *Arthritis and Rheumatism*, 59, 632–641. <https://doi.org/10.1002/art.23563>
- Hill, J. C., Whitehurst, D. G. T., Lewis, M., Bryan, S., Dunn, K. M., Foster, N. E., Konstantinou, K., Main, C. J., Mason, E., Somerville, S., Sowden, G., Vohora, K., & Hay, E. M. (2011). Comparison of stratified primary care management for low back pain with current best practice (STarT Back): A randomised controlled trial. *Lancet*, 378, 1560–1571. [https://doi.org/10.1016/S0140-6736\(11\)60937-9](https://doi.org/10.1016/S0140-6736(11)60937-9)
- Hjollund, N. H., Larsen, F. B., & Andersen, J. H. (2007). Register-based follow-up of social benefits and other transfer payments: Accuracy and degree of completeness in a Danish interdepartmental administrative database compared with a population-based survey. *Scandinavian Journal of Public Health*, 35, 497–502. <https://doi.org/10.1080/14034940701271882>
- Hodges, P. W. (2019). Hybrid approach to treatment tailoring for low back pain: A proposed model of care. *The Journal of Orthopaedic and Sports Physical Therapy*, 49, 453–463. <https://doi.org/10.2519/jospt.2019.8774>
- Hsu, C., Evers, S., Balderson, B. H., Sherman, K. J., Foster, N. E., Estlin, K., Levine, M., & Cherkin, D. (2019). Adaptation and Implementation of the STarT back risk stratification strategy in a US health care organization: A process evaluation. *Pain Medicine*, 20(6), 1105–1119. <https://doi.org/10.1093/pm/pny170>
- Janssen, M. F., Pickard, A. S., Golicki, D., Gudex, C., Niewada, M., Scalone, L., Swinburn, P., & Busschbach, J. (2013). Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: A multi-country study. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*, 22, 1717–1727. <https://doi.org/10.1007/s11136-012-0322-4>
- Kent, P. M., Keating, J. L., & Taylor, N. F. (2009). Primary care clinicians use variable methods to assess acute nonspecific low back pain and usually focus on impairments. *Manual Therapy*, 14, 88–100. <https://doi.org/10.1016/j.math.2007.12.006>
- Khan, Y., Lawrence, D., Vining, R., & Derby, D. (2019). Measuring biopsychosocial risk for back pain disability in chiropractic patients using the STarT back screening tool: A cross-sectional survey. *Chiropractic & Manual Therapies*, 27, 2. <https://doi.org/10.1186/s12998-018-0228-5>
- Kongsted, A., Andersen, C. H., Hansen, M. M., & Hestbaek, L. (2016). Prediction of outcome in patients with low back pain—A prospective cohort study comparing clinicians' predictions with those of the start back tool. *Manual Therapy*, 21, 120–127. <https://doi.org/10.1016/j.math.2015.06.008>
- Konstantinou, K., Dunn, K. M., Ogollah, R., Lewis, M., van der Windt, D., & Hay, E. M. (2018). Prognosis of sciatica and back-related leg pain in primary care: The ATLAS cohort. *The Spine Journal: Official Journal of the North American Spine Society*, 18, 1030–1040. <https://doi.org/10.1016/j.spinee.2017.10.071>
- Main, C. J., Sowden, G., Hill, J. C., Watson, P. J., & Hay, E. M. (2012). Integrating physical and psychological approaches to treatment in low back pain: The development and content of the STarT Back trial's 'high-risk' intervention (STarT Back; ISRCTN 37113406). *Physiotherapy*, 98, 110–116. <https://doi.org/10.1016/j.physio.2011.03.003>
- Medeiros, F. C., Costa, L. O. P., Oliveira, I. S., Oshima, R. K., & Costa, L. C. M. (2018). The use of STarT BACK Screening Tool in emergency departments for patients with acute low back pain: A prospective inception cohort study. *European Spine Journal*, 27, 2823–2830. <https://doi.org/10.1007/s00586-018-5586-0>
- Moher, D., Schulz, K. F., & Altman, D. (2005). The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomized trials 2001. *Explore*, 1(1), 40–45. <https://doi.org/10.1016/j.explore.2004.11.001>
- Molgaard Nielsen, A., Hestbaek, L., Vach, W., Kent, P., & Kongsted, A. (2017). Latent class analysis derived subgroups of low back pain patients—do they have prognostic capacity? *BMC Musculoskeletal Disorders*, 18, 345. <https://doi.org/10.1186/s12891-017-1708-9>
- Morso, L., Albert, H., Kent, P., Manniche, C., & Hill, J. (2011). Translation and discriminative validation of the STarT back screening tool into Danish. *European Spine Journal*, 20, 2166–2173. <https://doi.org/10.1007/s00586-011-1911-6>
- Morso, L., Kent, P., Albert, H. B., Hill, J. C., Kongsted, A., & Manniche, C. (2013). The predictive and external validity of the STarT back tool in Danish primary care. *European Spine Journal*, 22, 1859–1867. <https://doi.org/10.1007/s00586-013-2690-z>
- Morso, L., Kongsted, A., Hestbaek, L., & Kent, P. (2016). The prognostic ability of the STarT back tool was affected by episode duration. *European Spine Journal*, 25, 936–944. <https://doi.org/10.1007/s00586-015-3915-0>

- Morso, L., Schiøttz-Christensen, B., Søndergaard, J., Andersen, N.-B., Pedersen, F., Olsen, K. R., Jensen, M. S., Hill, J., & Christiansen, D. H. (2018). The effectiveness of a stratified care model for non-specific low back pain in Danish primary care compared to current practice: Study protocol of a randomised controlled trial. *Trials*, *19*, 315. <https://doi.org/10.1186/s13063-018-2685-5>
- Oliveira, C. B., Pinheiro, M. B., Teixeira, R. J., Franco, M. R., Silva, F. G., Hisamatsu, T. M., Ferreira, P. H., & Pinto, R. Z. (2019). Physical activity as a prognostic factor of pain intensity and disability in patients with low back pain: A systematic review. *European Journal of Pain*, *23*, 1251–1263. <https://doi.org/10.1002/ejp.1395>
- Regional management program for low back pain, the Region of Southern Denmark. (2010).
- Riis, A., Jensen, C. E., Bro, F., Maindal, H. T., Petersen, K. D., Bendtsen, M. D., & Jensen, M. B. (2016). A multifaceted implementation strategy versus passive implementation of low back pain guidelines in general practice: A cluster randomised controlled trial. *Implementation Science*, *11*, 143. <https://doi.org/10.1186/s13012-016-0509-0>
- Riis, A., Rathleff, M. S., Jensen, C. E., & Jensen, M. B. (2017). Predictive ability of the start back tool: An ancillary analysis of a low back pain trial from Danish general practice. *BMC Musculoskeletal Disorders*, *18*, 360. <https://doi.org/10.1186/s12891-017-1727-6>
- Roland, M., & Fairbank, J. (2000). The Roland–Morris disability questionnaire and the Oswestry disability questionnaire. *Spine*, *25*(24), 3115–3124. <https://doi.org/10.1097/00007632-200012150-00006>
- Savigny, P., Kuntze, S., Watson, P., Underwood, M., Ritchie, G., Cotterell, M., Walsh, D. (2009). *Low back pain: Early management of persistent non-specific low back pain*. National Collaborating Centre for Primary Care and Royal College of General Practitioners.
- Sowden, G., Hill, J. C., Konstantinou, K., Khanna, M., Main, C. J., Salmon, P., Somerville, S., Wathall, S., & Foster, N. E. (2012). Subgrouping for targeted treatment in primary care for low back pain: The treatment system and clinical training programmes used in the IMPaCT Back study (ISRCTN 55174281). *Family Practice*, *29*(1), 50–62. <https://doi.org/10.1093/fampra/cmr037>
- Stochkendahl, M. J., Kjaer, P., Hartvigsen, J., Kongsted, A., Aaboe, J., Andersen, M., Andersen, M. Ø., Fournier, G., Højgaard, B., Jensen, M. B., Jensen, L. D., Karbo, T., Kirkeskov, L., Melbye, M., Morsel-Carlsen, L., Nordsteen, J., Palsson, T. S., Rasti, Z., Silbye, P. F., ... Vaagholt, M. (2018). National clinical guidelines for non-surgical treatment of patients with recent onset low back pain or lumbar radiculopathy. *European Spine Journal*, *27*, 60–75. <https://doi.org/10.1007/s00586-017-5099-2>
- Trinderup, J. S., Fisker, A., Juhl, C. B., & Petersen, T. (2018). Fear avoidance beliefs as a predictor for long-term sick leave, disability and pain in patients with chronic low back pain. *BMC Musculoskeletal Disorders*, *19*, 431. <https://doi.org/10.1186/s12891-018-2351-9>
- Twisk, J. (2013). *Applied Longitudinal Data Analysis for Epidemiology: A Practical Guide*. Cambridge University Press.
- van der Heijden, G. J., Donders, A. R., Stijnen, T., & Moons, K. G. (2006). Imputation of missing values is superior to complete case analysis and the missing-indicator method in multivariable diagnostic research: A clinical example. *Journal of Clinical Epidemiology*, *59*, 1102–1109. <https://doi.org/10.1016/j.jclinepi.2006.01.015>
- Werneke, M. W., Edmond, S., Young, M., Grigsby, D., McClenahan, B., & McGill, T. (2020). Association between changes in function among patients with lumbar impairments classified according to the STarT back screening tool and managed by McKenzie credentialed physiotherapists. *Physiotherapy Theory and Practice*, *36*, 589–597. <https://doi.org/10.1080/09593985.2018.1490839>

## SUPPORTING INFORMATION

Additional Supporting Information may be found online in the Supporting Information section.

**How to cite this article:** Morsø, L., Olsen Rose, K., Schiøttz-Christensen, B., Sowden, G., Søndergaard, J., & Christiansen, D. H. (2021). Effectiveness of stratified treatment for back pain in Danish primary care: A randomized controlled trial. *European Journal of Pain*, *25*, 2020–2038. <https://doi.org/10.1002/ejp.1818>