Effectiveness and mechanisms of a multimodal treatment for low back pain: a pragmatic matched cohort study

ROBIN SCHÄFER[1,2], DANIEL NIEDERER[3], CLAUDIA LEVENIG[4], MONIKA HASENBRING[4], THOMAS TAS[1,5], DANIELA FETT[1], KATHARINA TROMPETER[1,2], THORE HAAG[6], CHRISTIAN SCHNEIDER[6], PHILIPP FLOESSEL[7], HEIDRUN BECK[7], MARCUS SCHILTENWOLF[8], TILMAN ENGEL[9], FRANK MAYER[9], PETRA PLATEN[1]

[1] *Department of Sports Medicine and Sports Nutrition, Faculty of Sports Science, Ruhr University Bochum*
[2] *Division of Physiotherapy, Department of Applied Health Sciences, University of Applied Sciences, Bochum*
[3] *Department of Sports Medicine and Exercise Physiology, Institute of Occupational, Social and Environmental Medicine, Goethe University Frankfurt*
[4] *Department of Medical Psychology and Medical Sociology, Faculty of Medicine, Ruhr University of Bochum*
[5] *Novotergum GmbH*
[6] *Schön Klinik München Harlaching, Sportorthopädisches Institut*
[7] *UniversitätsCentrum für Orthopädie, Unfall- & Plastische Chirurgie, Universitätsklinikum Carl Gustav Carus Dresden*
[8] *Klinik für Orthopädie und Unfallchirurgie, Konservative Orthopädie und Schmerztherapie, Universitätsklinikum Heidelberg*
[9] *University Outpatient Clinic, Sports Medicine & Sports Orthopaedics, University of Potsdam*

*** Correspondence:** robin.schaefer@rub.de

Cite as: Schäfer, R., Niederer, D., Levenig, C., Hasenbring, M., Tas, T., Fett, D., Trompeter, K., Haag, T., Schneider, C., Flößel, P., Beck, H., Schiltenwolf, M., Engel, T., Mayer, F., Platen, P. (2022). Effectiveness and mechanisms of a multimodal treatment for low back pain: a pragmatic matched cohort study. *MedRxiv.*

Supplemental material: https://osf.io/mbrh9/

Abstract

Objective: To investigate the effectiveness and mechanisms of a multimodal treatment including perturbation exercise directly applied in health care.

Methods: Participants from the intervention group had chronic or recurrent low back pain and participated in a 12-week back pain prevention program. Coaches were trained to deliver multimodal care. Controls (usual care) were matched from a multi-center RCT. Outcomes were pain, disability, isokinetic trunk strength (extension/flexion) and balance (center of pressure trace). A Bayesian, sequential analysis along 8 matching procedures and moderation/mediation analyses incorporating the biopsychological avoidance-endurance model were conducted. Median values with highest posterior density intervals (HPDI) from baseline-adjusted analyses are presented.

Results: 128 participants were matched to control. Over 12 weeks, intervention and control (n = 128 each) experienced a similar decrease in pain and disability, which led to negligible average treatment effects for pain (0.5 $HPDI_{95\%}$[-3.6,4.6]) and disability (1.5 $HPDI_{95\%}$[-2.3,5.4]). Changes in functional parameters (n=18) showed small effects in favor of the intervention group, in particular for trunk extension (4.1 Nm $HPDI_{95\%}$[-10,18.2]) and monopedal stances (standardized score: 0.49 $HPDI_{95\%}$ [0.13,0.79]). Depression was higher in drop-outs and decreases in pain/disability were associated with decreases in depression. Distress-endurance subgroups experienced higher baseline pain and disability and showed the highest reductions in both parameters upon completion of the intervention.

Conclusion: Multimodal treatments for low back pain without tailoring are possibly less effective in the context of health care than their efficacy in RCTs suggests. Targeting distress-endurance subgroups with a multimodal treatment approach is probably an effective strategy in treatment tailoring.

Keywords: avoidance-endurance, sensorimotor exercise, perturbation, biopsychosocial, moderation, mediation, treatment mechanism

Introduction

The burden of low back pain (LBP) is high [8] and treatment effect sizes are usually only small to medium [18]. One reason for that low effect is that primary pathologies and optimal treatments for relevant subgroups in the LBP domain can barely be found [33]. Therefore, the use of a biopsychosocial model for explaining LBP (especially at the chronic stage or chronification) and guiding treatment (e.g. multidisciplinary/-modal) is emphasized [4, 19]. While education and exercise are almost consistently recommended as first-line treatment [12], the use of a costly and time-intensive multidisciplinary approach should be tailored (e.g. for chronic LBP with significant psychosocial contribution) [19].

The biopsychosocial avoidance-endurance model from Hasenbring et al. [13] aims to describe the chronification of pain by two maladaptive patterns: *avoidance* and *endurance* (*overdoing)* [6]. Endurance-related responses can be further distinguished to distress-endurance (pain-related thought suppression, depressive mood) and eustress-endurance (focused distraction, positive mood) which both result in task-persistence and physical overuse [16]. The fear-avoidance response includes pain-related avoidance of activity and catastrophizing. The fear-avoidance and endurance response result in muscular disuse and therefore negatively affect pain intensity, disability, pain severity and quality of life [23]. Consequently, *activity pacing* (adaptive response) is proposed as an optimal reaction to the onset of pain. High levels of activity pacing and low levels of avoidance were associated with better outcomes for physical functioning [6]. The distress-endurance might be the most critical subgroup regarding experienced pain and disability [11, 14, 36]. However, we do not know yet if those subgroups react differently to a multimodal treatment including exercise. Further, it is unknown whether changes in underlying subscales according to the avoidance-endurance model mediate changes in pain and disability outcomes.

In this study, we aim to examine the *effectiveness* of a multimodal treatment (sensorimotor exercise with perturbation + education) directly integrated into health care as a *pragmatic trial* [7, 10, 31]. The effects are compared to a matched control group from a multi-center RCT, which will evaluate the exercise component's *efficacy* [26, 27]. Further, we want to examine moderating (i.e., for whom is the treatment effective?) and mediating (i.e, how does the treatment work?) factors by integrating the avoidance-endurance model. Specifically, we hypothesize that changes towards adaptive behavior lead to beneficial changes in pain and disability.

Methods

Study Design and Ethics

We carried out a pragmatic, prospective cohort study with matched pairs. The study was conducted in agreement with the declaration of Helsinki, approved by the local ethics committee of the Faculty of Sport Science of Ruhr-University Bochum (EKS V 10/2017) and participants gave their written informed consent. The trial is registered in the German Clinical Trial Register (DRKS00030389). The study report follows the guidelines CERT [32] and the CONSORT extension for pragmatic trials [40].

Participants

Intervention group

Participants were recruited from a specific back pain prevention program. The program was offered by their health insurance company (AOK Nordwest) free of charge. Participants typically suffered from chronic or recurrent low back pain. They either self-registered for the program or were offered this program in a personal health consultation by employees of the health insurance company or the executing coaches.

Control group & matching

The control group was matched from a cohort of a multi-center study conducted within our study network [27]. The control group received usual care. However, they also received a more frequent and extensive diagnostic assessment than the intervention group within the given timeframe of 12 weeks. The matching was done by calculating the (multivariate) Mahalanobis Distance between each observation and greedy nearest neighbor matching. To increase sensitivity, 8 different matching variable sets were used to receive 8 possibly overlapping control groups. The variables pain, disability, age, gender, height, weight, body mass index and chronification at baseline in different combinations were used, where gender was matched exactly. The matching procedure was carried out in Matlab (2020a). The code and more detailed information are available in our OSF repository (https://osf.io/mbrh9/).

Inclusion criteria

Inclusion criteria for both groups were: age 18-65, at least one episode (≥ 4 days) of nonspecific back pain in the last 12 months, ability to understand the extent and meaning of the study and to answer a questionnaire without help [27].

Intervention

The intervention consisted of a 12-week multimodal group intervention program including sensorimotor exercises with perturbations, and theoretical education. The intervention took place at multiple locations and was, thus, given by different coaches. Prior to the study, all coaches received a full-day workshop and were educated in theory and practice about the content of the intervention program by a multidisciplinary team of our research network.

The exercise components comprised four core exercises (plank, side-plank, single-leg stance, rowing) which were varied in stability and strength demands [27, 30]. The participants of the course were supposed to exercise one session per week á 60 minutes in the group supervised by the coaches (center-based) and were instructed to do two additional weekly sessions á 30 minutes at home (home-based). The home-based training was not documented by the participants or the coaches.

The education components as part of the center-based sessions included pain neurophysiology, pain coping strategies according to the avoidance-endurance model [14, 16] and other relevant topics in the context of back pain (e.g. physical activity, activity pacing, stress & relaxation, risk factors, sensorimotor exercise).

According to the CERT guideline, further details are presented in the supplement (https://osf.io/c84rt).

Outcomes

The main outcomes were pain and disability. A subset of the intervention group was tested for isokinetic trunk strength (extension/flexion) and balance (postural sway). Also, changes in subscales related to the avoidance-endurance classification [15] (e.g., depression, pain persistence) as well as discrete group changes based on these parameters were evaluated. Further, the moderating and mediating effects of avoidance-endurance parameters on pain and disability were analyzed.

Pain & Disability

The 7-item Chronic Pain Grade (CPG) questionnaire by von Korff et al [20] was used to assess pain intensity, disability and grade of chronification. The mean of items 1-3 and 4-6 built the subscales for pain intensity and disability (scale: 0-100), respectively. Utilizing item 7 together with these subscales, the grade of chronification was calculated as described in the original literature [20].

Strength & Balance

The maximum isokinetic trunk strength (extension and flexion) was tested with the IsoMed 2000 Back Module (D&R GmbH, Germany). After a warm-up phase of 30 submaximal repetitions, the test consisted of five maximum repetitions of 20° extension and 30° flexion in a sitting position. The movement speed was constant at 60°/s. The average value of the three highest repetitions was used as outcome [Nm] for flexion and extension.

Balance was tested in bipedal and monopedal stances (left and right) over 30 s. The trace of the center of pressure [mm] was calculated as an outcome parameter, where the sum of both monopedal stances was used as outcome. Stances were standardized (hands gripping the hip, barefoot, visual fixation in 2-meter distance). The tests were performed on different force platforms for the intervention (Balance Trainer BTG4, Hur Labs, Finnland) and the control group (CSMi Computer Sports Medicine Inc., Stoughton, MA, USA). We performed a direct comparison of these devices and both instruments yielded valid results for the postural sway [3]. However, to account for heterogeneity, pre and post scores were standardized by the mean and standard deviation of the pre-test within each group (intervention/control).

Depression

Depression was measured via the German version of the Beck Depression Inventory for Primary Care (BDI-PC) [28] and two extra items ("loss of pleasure" and "incapacity to decide") from the long BDI [2, 21]. The BDI-PC consists of 7 items, where the sum of those builds the subscale for depression ranging from 0 (no depression) to 21 (highest possible depression).

Avoidance-Endurance

Psychosocial responses were measured via the Avoidance-Endurance Questionnaire (AEQ). The questionnaire consists of 28 items. The pain persistence and humor/distraction subscales are part of the behavioral endurance scale of the AEQ. Further scales used in this study were thought suppression and hopelessness.

Subgroup classification

The classification of the avoidance-endurance subgroups was based on 1) the pain persistence scale (7 items) from the AEQ and the BDI-PC (7 items) and 2) the pain persistence scale and two items from the extended BDI version [36, 39]. In both procedures, pain persistence values equal or higher than 3 led to an endurance classification (either distress- or eustress-endurance) and otherwise to adaptive/avoidance response. For 1), depression values equal to or higher than 2 led to a distress-endurance or fear-avoidance classification and otherwise to an eustress-endurance or adaptive classification. For 2), the same classification was applied if both depression items were rated 1 or higher [41].

Statistical Analyses

The results should be considered rather explorative and hypothesis-generating than confirmative. Therefore, no significance testing was performed. Robustness was checked in several sensitivity analyses. The effects were described using *confidence* or *highest posterior density intervals*, which we aim to interpret as compatibility intervals [1]. All analyses were conducted in R [29].

Treatment effects

We conducted a Bayesian, sequential analysis [9] using a simple change score and a baseline-adjusted model without additional covariates for the outcomes pain and disability. This was done for all 8 matching scenarios. The model was updated with each matched pair (n=126) for each outcome. This yields $2_{(outcomes)} * 2_{(models)} * 8_{(matchings)} * 126_{(matched\ pairs)} = 4032$ model fits. The models were fitted using

the brms package [5] using a weakly informative prior for the outcome (normally distributed change score with mean = 0 and standard deviation = 10). The *highest posterior density intervals* (e. g. 95%) from the posterior distribution of the treatment effect are visualized and reported. Strength and balance outcomes were analyzed for all 8 matching procedures yielding $4_{(outcomes)} * 8_{(matchings)} = 32$ (baseline-adjusted) models. Further, we analyzed discrete changes in *avoidance-endurance* classifications and changes in the underlying subscales mostly on a descriptive basis.

Moderation

Moderation (effect measure modification) analyses for the outcomes pain and disability were performed for avoidance-endurance subgroups via simple (frequentist) linear regression with the change of pain and disability as outcome and subgroup as a predictor. Baseline-adjusted models were used for sensitivity analysis.

Mediation

We further evaluated if the treatment works via changes in subscales of the avoidance-endurance model affecting pain and disability as outcomes. The within-participant mediation model from Montoya & Hayes [24] was adopted and enhanced to multiple mediators (depression, hopelessness, thought suppression, pain persistence, distraction) and multiple outcomes (pain, disability) (Figure 1). Thus, direct and indirect (mediated) treatment effects were calculated for each subscale and each outcome. The grand mean centered parametrization from Montoya and Hayes [24] was used to meaningfully interpret the results. The models were fitted via the Bayesian regression package brms [5]. The full posterior distribution is presented for direct, indirect and total effects.

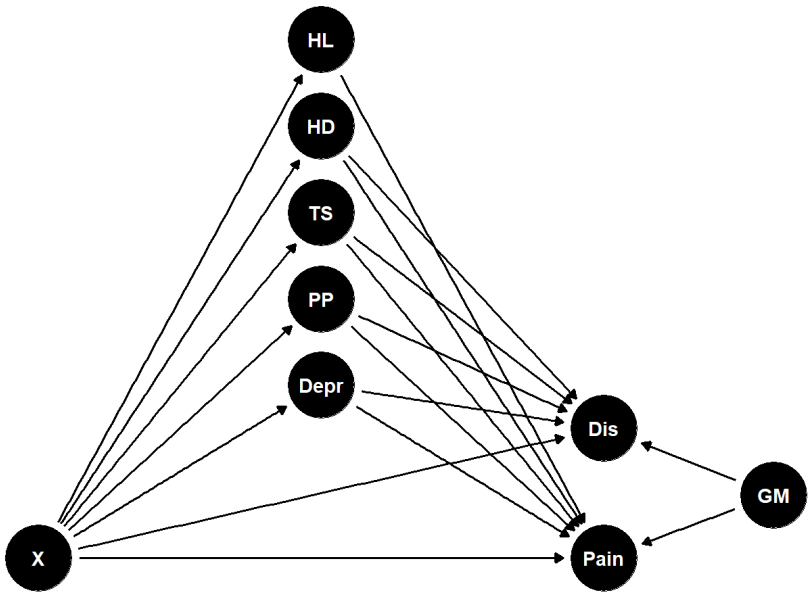

Figure 1. Causal Graph of the one-condition within-participant mediation analysis with multiple serial mediators and 2 outcomes adapted from Montoya & Hayes [33]. The nodes for mediators and outcomes represent pre-post differences. X = treatment (intervention only), Mediators: Depr = depression, PP = pain persistence, TS = thought suppression, HD = humor/distraction, HL = hopelessness. Outcomes: Dis = disability, Pain = pain intensity, Parametrization: GM = grand mean

Results

Matching procedure and subjects characteristics

Figure 2 shows the participant flow for the 8 matching procedures. Table 1 shows covariate balance before and after matching for all participants and the subset that received functional tests (strength & balance). Dropout during the intervention period was high (60.1%, n = 190 who did not show up at the last course session or did not answer the questionnaire). The trainers reported no adverse events. The baseline values of participants who dropped out tended to be higher for psychological subscales (e.g., hopelessness, catastrophizing, thought suppression) and pain/disability (see our repository https://osf.io/cfnw5).

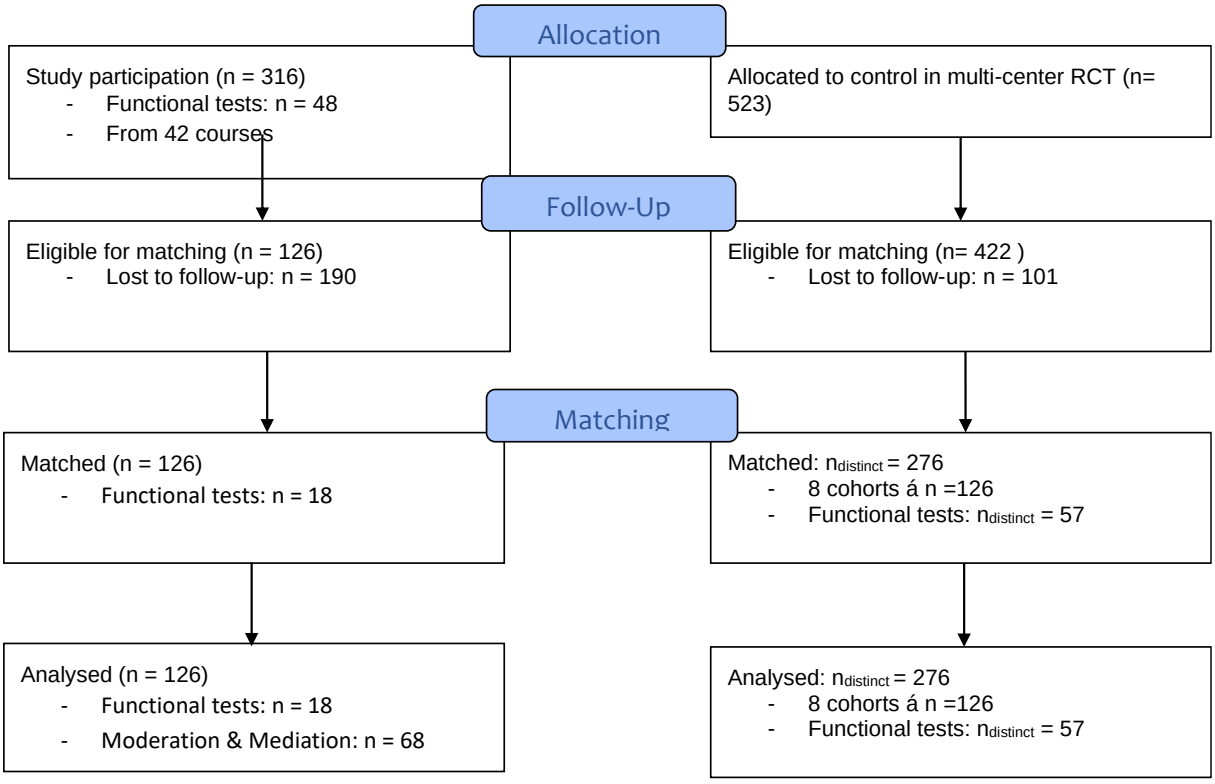

Figure 2: CONSORT flow diagram.

Table 1: Covariates of the intervention and control group before and after matching

Group	Outcome	n	Female N	age years	Height cm	Weight kg	Pain [0-100]	Disability [0-100]
Intervention	Pain/Disability	126	88	50.9 (13.8)	170.5 (8.2)	72.8 (13.4)	42.5 (21.6)	25.5 (25)
Intervention	Strength/Balance	18	10	47.3 (13.6)	174.4 (8.7)	78.3 (13.5)	42.6 (20.6)	18.3 (14.7)
Control - before		422	234	40.0 (13.9)	173.6 (9.0)	74.7 (15.2)	32.6 (17.3)	19.2 (19.8)
Control - after	Pain/Disability	276	184	43.6 (13.9)	171.9 (8.7)	73.5 (14.7)	36.6 (17.3)	21.8 (20.8)
Control - after	Strength/Balance	57	34	46.1 (13.3)	172.1 (9.3)	76.2 (14.2)	37.7 (15.9)	16.2 (13.1)

Pain & Disability

The average decrease in both groups was similar for pain (intervention: 8.0 SD: 18.7, control: 5.8-8.6 SD: 16.6-18.4) and disability (intervention: 6.5 SD: 21.5, control: 6.1-7.4 SD: 18.6-20.4) The baseline-adjusted treatment effect compared to control over all matching procedures was 0.5 (median) $HPDI_{95}$ [-3.6; 4.6] for pain and 1.5 $HPDI_{95\%}$ [-2.3; 5.4] for disability. The difference in change scores yielded similar results for pain (-0.9 $HPDI_{95\%}$ [-5.5; 3.7] and disability (0.2 $HPDI_{95\%}$ [-4.7; 5.1]) – see Figure 3. Negative values favor the intervention group. The sequential analysis shows no substantial influence on the overall matching procedure or the 8 separate matchings (for details see our repository https://osf.io/mbrh9/).

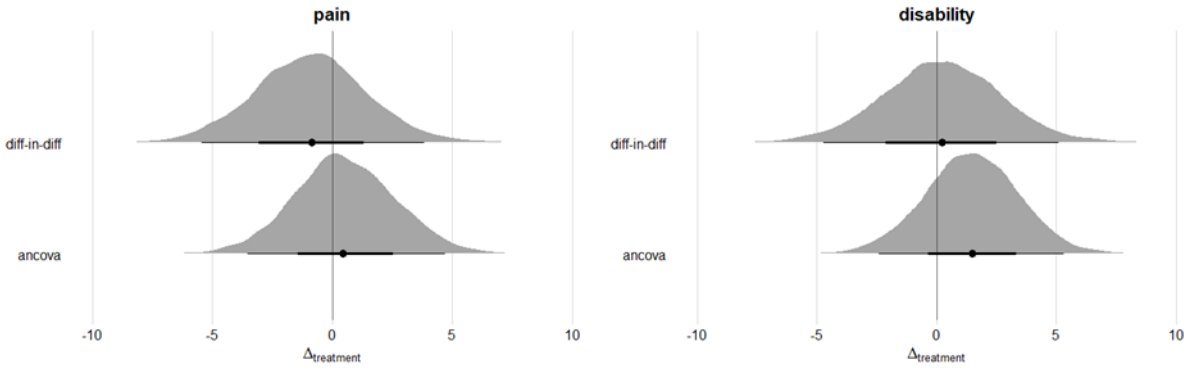

Figure 3. Posterior distributions for the average treatment effect (x-axis) compared to control for pain and disability for simple change score (*diff.in.diff*) and baseline-adjusted analysis (*ANCOVA*) – n = 126. These distributions are averaged over all 8 matching procedures. Dots show the median of the posterior distribution and bold/thin whiskers represent 95/50% HPDI (highest posterior density interval)

Strength & Balance

The baseline-adjusted analyses compared to control show mostly favorable effects for the treatment group (see Figure 4, positive values favor the intervention group). Postural sway reductions were lower for intervention in monopedal (sum of left and right; z = 0.49 $HPDI_{95\%}$ [0.13, 0.79]) compared to bipedal stances (0.19 $HPDI_{95\%}$ [-0.29, 0.71]). Strength increase was higher for intervention in trunk extension (4.1 Nm $HPDI_{95\%}$ -10.0, 18.2]) and flexion (11.1 Nm $HPDI_{95\%}$ [2.4, 19.4]).The treatment effects were more uncertain (lower and wider curves) and smaller for the bipedal stance and trunk extension.

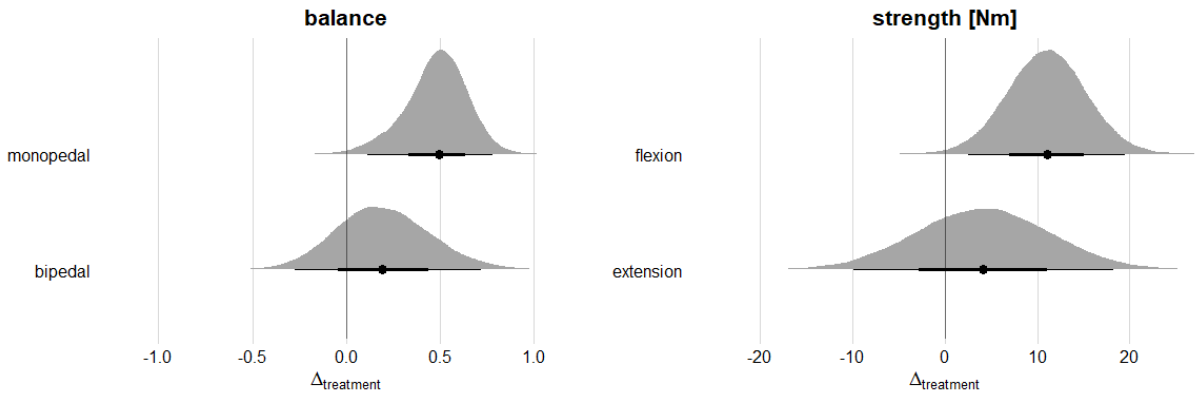

Figure 4. Posterior distributions for the baseline-adjusted (ANCOVA) average treatment effects (x-axis) for the functional outcomes balance and strength – n = 18. These distributions are averaged over all 8 matching procedures. Dots show the median of the posterior distribution and bold/thin whiskers represent 95/50% HPDI (highest posterior density interval)

Avoidance-Endurance Parameters

In the intervention group (n=68), the mean reductions for pain persistence (scale: 0 to 6) and depression (scale: 0 to 9) were 0.1 (SD: 0.9) and 0.9 (SD: 2.4), respectively. These changes were driven by the participants who had high baseline values. Given that the subgroups were determined by depression and pain persistence, this led to substantial positive changes in subgroup classifications after the treatment (see Table 2). The favorable adaptive classification increased from 16 to 30 and the maladaptive distress-endurance type was reduced from 22 to 12.

Table 2. Avoidance-Endurance classifications before and after the treatment in the intervention group (n=68). The classification is based on *pain persistence* and *depression* subscales.

	Adaptive	Distress-Endurance	Eustress-Endurance	Fear-Avoidance
Pre	16	22	20	10
Post	30	12	19	7

Moderation - for whom is the treatment effective?

Evaluating the treatment group only (n=68), pain and disability reductions were highest for the distress endurance subgroup (see Table 3). However, adjusting the analysis for baseline values diminished the outcome for distress-endurance. In this subsample, reductions were slightly higher for pain (-9.9 SD: 17.5) and disability (-9 SD: 23.4) compared to the complete matched cohort (n=126).

Table 3. Moderation analysis for pain and disability reductions stratified by avoidance-endurance subgroups at baseline in the intervention group (n=68)

		Adaptive		Distress-Endurance		Eustress-Endurance		Fear-Avoidance	
Outcome	Analysis	Mean	$CI_{95\%}$	Mean	$CI_{95\%}$	Mean	$CI_{95\%}$	Mean	$CI_{95\%}$
Pain	Delta	4.8	[-2.9, 12.4]	14	[6.9, 21.1]	10.7	[2.5, 19]	9.4	[-1.1, 19.9]
Pain	ANCOVA	8.6	[1.4, 15.7]	11.8	[5.2, 18.4]	9.6	[2.1, 17.1]	8.9	[-0.7, 18.5]
Disability	Delta	5.1	[-5.2, 15.4]	12.6	[3, 22.3]	10.2	[-0.9, 21.3]	7	[-7.3, 21.2]
Disability	ANCOVA	9.9	[0.1, 19.7]	9.8	[0.8, 18.8]	8.7	[-1.5, 19]	6.4	[-6.7, 19.4]

Stratification by CPG grades shows, that the low pain group (grade 1, n = 74) did not improve on average in either pain intensity (2.0 $CI_{95\%}$ [-1.7, 5.8]) or disability (-0.4 $CI_{95\%}$ [-4.8, 3.9]). The longitudinal change was, thus, largely driven by the intermediate/high pain groups (grade 2, n = 39) with effect sizes of 18.2 $CI_{95\%}$ [13.1, 23.3] for pain and 13.4 $CI_{95\%}$ [7.5, 19.4] for disability.

Mediation – how does the treatment work?

Within the intervention group, changes in depression and hopelessness seem to be substantially associated with the changes in pain and disability (Figure 5). The indirect effect of depression on pain was -2.2 $HPDI_{90\%}$ [-4.4, -0.6] and on disability -1.9 $HPDI_{90\%}$ [-4.3 -0.1]. The sum of indirect effects was -3.5 $HPDI_{90\%}$ [-6.4, -0.8] for pain and -4.7 $HPDI_{90\%}$ [-8.6, -0.9] for disability, indicating potential treatment

mechanisms by the supposed mediators The total proportion mediated was 34.1 % for pain and 49.3 % for disability.

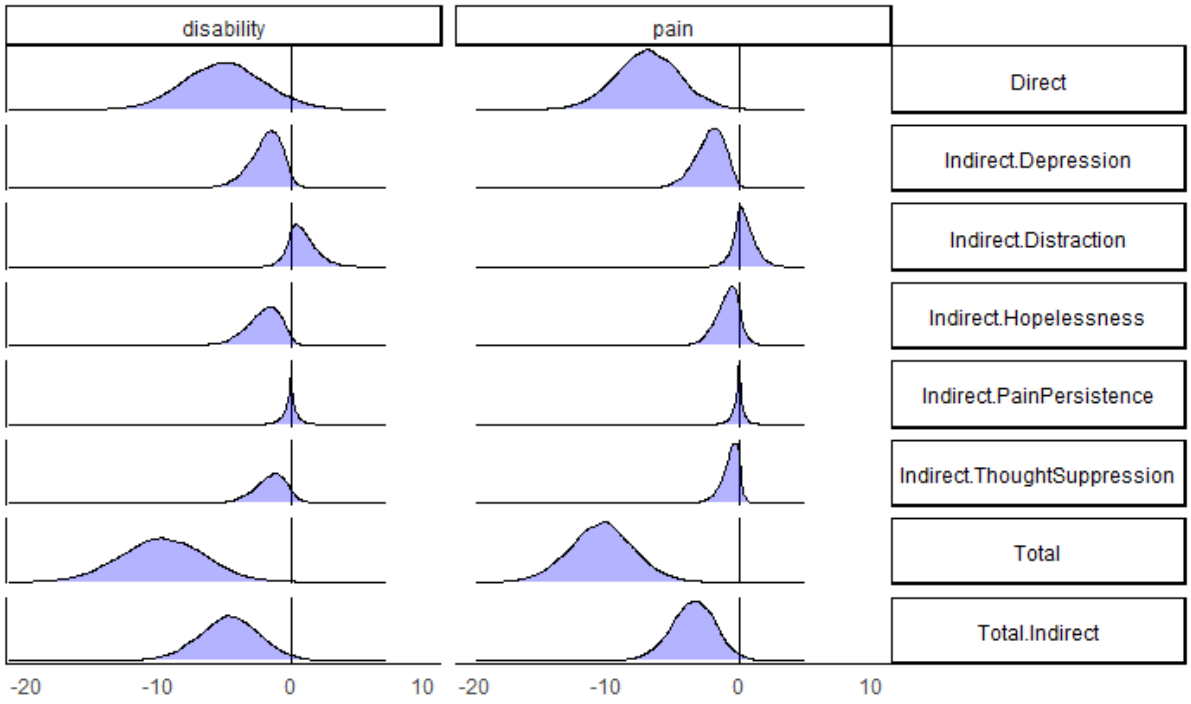

Figure 5. Posterior distributions for the direct, indirect and total (indirect) effects of the within-participant, (serial) multiple mediation model with two outcomes (pain/disability).

Discussion

We found similar decreases in the multimodal intervention group and the matched control groups for pain and disability. However, small treatment effects for strength and balance were in favor of the intervention. Distress-endurance subgroups showed the highest pain and disability scores and also the highest reductions upon completion of the intervention.

Population and sample representativeness

A substantial part of the study population experienced chronic pain and maladaptive pain responses including depression symptoms. The classification approach via pain persistence and the BDI-PC yielded 32% distress-endurance responses, which is between study results from a subacute (19%) [14] and a chronic population (34 %) [11]. Additionally. About one third of the intervention group showed high grades of chronification (Korff grades 2 and 3) with high pain and disability. Those people would rather seek care in the rehabilitative or tertiary care sector.

Nevertheless, the concept of the health care program we evaluated in this study aims at prevention, anchored in German social law (SGB V $20a). In our experience, this incongruency is common in practice, as minimal exclusion criteria exist. Consequently, practitioners are challenged by heterogenous groups, comprising the overall quality of such programs. The avoidance-endurance model and other classification systems [18] might help to assign adequate treatments.

Pain intensity & Disability

When comparing treatment to matched control, the finding of a "null" effect for pain and disability might surprise. In fact, both groups experienced reductions in pain and disability. One may suspect that the effects of an intervention applied directly in health care might yield on average inferior results to a strictly controlled RCT [10]. Examples of such discrepancies between *efficacy* and *effectiveness in* the health care sector are already given by Cochrane (1972) [7] and Treweek & Zwarenstein [37]. The disability reduction in sensorimotor exercise under controlled conditions was on average higher compared to our results (10.2 vs 6.5 points), while the control group also reduced disability by a fair amount (4.9 points) [26]. This "placebo" effect can be partly explained by contextual factors as part of a strictly controlled RCT, for example, visiting the laboratory several times and reminders to increase study adherence [31], but may also arise from selection bias (e.g. participants of back pain studies are likely to suffer recently from back pain and get better soon without treatment).

Strength & postural control

The treatment effects for strength and postural control were surprisingly high. Unpublished results from a multi-center RCT [27] on sensorimotor exercise only show rather small effect sizes ($d \approx 0.2$) for extension and monopedal stance and rather negligible effects for bipedal stance and flexion in a between-group comparison. In our study, the comparison with a matched control group from the same study yielded larger effect sizes for balance. The effects for flexion were higher and for extension small and uncertain. We suppose, that this can be partly explained by stronger balance and strength deficits in the intervention group indicated by higher age even after matching. For such people, the sensorimotor exercise may provide a much higher stimulus for adaptation in balance and strength including fast adaptations in the early phase of training. This argument is even stronger for strength adaptations, as sensorimotor exercise does not particularly aim to increase maximal strength. However, low baseline values also may yield larger effect sizes due to regression to the mean [34].

Clinical relevance: For whom is the treatment effective?

While considering that average effect sizes for non-specific pain are relatively small [18] in controlled studies, we conclude that our study provides evidence, that the average treatment effect for a multimodal LBP intervention is substantially lower when applied directly in health care settings. This highlights the importance of 1) evaluating the dissemination process of treatments (e.g. effectiveness trials) and 2) peeling off the “non-specific” label and, hence, identifying subgroups and treatment mechanisms to achieve more effective treatment strategies to ameliorate the burden of low back pain [17, 18, 33]. Given this, we question the meaning of an average treatment effect for non-specific LBP and emphasize the statement from Hides et al. [17]: “The wrong question is to ask which approach is most effective”, but rather which treatment for whom is more effective – which in turn can be addressed by subgrouping and moderation analysis.

The multimodal treatment might be most effective for distress-endurance response types, which are classified by high depression and high pain persistence scales. This result is in line with Cane et al. [6], who suggested, that changes towards adaptive responses (activity pacing) yield functional outcomes. A reduction in distress-endurance responses from pre to post was also observed in our study (22 to 12). In their Cochrane review, Kamper et al. [19] found medium effect sizes for multidisciplinary treatment in chronic low back pain populations for pain ($d=0.55$) and disability ($d=0.41$) reductions after 12 weeks of intervention. It seems reasonable to allocate multidisciplinary treatment to those with significant psychosocial impact [19].

Sensorimotor exercise might be particularly good for learning activity pacing to overcome the endurance pattern. This kind of exercise promotes physical activity by coordinative/balance training with trunk stability exercise with low demands on physical strength which was shown to be effective in reducing pain and disability [26]. This way, endurance response types learn effective regulation.

Further, the educational component fosters self-reflection which should enhance self-efficacy via a biopsychosocial pathway. The average reductions in pain persistence and depression in almost all subgroups except for the adaptive response types at baseline might also indicate the effectiveness for the distress-endurance response type. The fear-avoidance response type might also benefit from this kind of intervention. The semi-standardized progression with a gradually increasing level structure can help to build trust in a movement to overcome kinesophobia.

How does the treatment work?

We found an association between changes in mood scales and changes in the outcomes. Murillo et al. [25] found that depression is a commonly used mediator in mediation analysis. They also found that the mediated proportion for cognitive behavioral treatment did not exceed 20% for any mediator. The (total) mediated proportion in this study is about twice as large, which could be explained by the within-mediation model without a control group. Surprisingly, the smallest indirect effect was found for pain persistence. Pain persistence was a crucial scale for classifying AEM subgroups in several studies [14, 35, 39]. In conclusion, the average decrease in pain persistence seems to be a minor driver for changes in pain and disability than mood characteristics.

Contrary to this analysis, Liew et al. [22] suggested based on their data-driven analysis, that changes in depression might be affected by earlier changes in pain and disability. This switch between mediator and outcome yields the same model fit at least in a 3-parameter model (e.g., exposure, mediator, outcome). Therefore, based on only the data we cannot conclude whether pain and disability are reduced because of reduced depression/mood symptoms or the other way round. Further research is highly needed, as in general, treatment mechanisms for chronic pain are scarcely investigated and heterogeneous in most cases [25].

Limitations

Due to the characteristics of this study (pragmatic trial), the drop-out was high and even higher among participants with high depression scores. One reason for the high drop-out rate was, missing direct contact with participants, i.e. the questionnaires were handed over by intermediaries like the health insurance and the coaches. In addition, the matched control group received a more extensive assessment and participated in a follow-up at 3 weeks but the intervention group did not. Furthermore, the intervention was not strictly controlled, leading to heterogeneity in delivery and, consequently, participant adherence. This heterogeneity was directly observed as feedback from the coaches who conducted the intervention. To overcome some of these issues, we carried out an extensive procedure (sequential matching analysis with sensitivity analysis), but the validity of the comparison might be questioned in light of the mentioned contextual factors. Further, the control group could not be used for moderation analysis, so we could not test the 3-fold interaction (time x intervention x AEM) but only a 2-fold interaction (time x AEM). Given that the derived AEM subgroup classifications are “natural” occurring groups rather than randomized, the coefficients from the baseline-adjusted moderation analyses are also biased estimators and do not yield causal estimates due to regression to the mean [34, 38]. Further, the sample size for the functional outcomes (n=18) is considerably low which might have led to inflated effect sizes.

Conclusion

Multimodal treatments for low back pain are probably less effective than efficacious in the context of health care. This adds burden to the “curse” of non-specific LBP, which is that average treatment effects are only low to moderate. Targeting depressive-endurance subgroups with a multimodal treatment approach is probably an effective strategy to be considered in treatment tailoring. A fast screen tool with 9 items can be used by practitioners for subgroup classification. While efficacy trials

under controlled conditions are warranted and necessary, effectiveness trials including the dissemination process should be evaluated more often.

Funding

This study was conducted within the MiSpEx research network and funded by the German Federal Institute for Sport Science (BISp) [ZMVI1-080102A/11-18].

Declaration of interest

The authors declare no conflict of interest.

Data availability

Data of the treatment group, study documents and the analysis script are available in our online repository: https://osf.io/mbrh9/

Contributions

Conceptualization – RS, DF, TT, PP, Methodology – RS; DF, TT, MH, DN, Formal Analysis – RS, MH, Investigation – RS; TT, Resources – RS; PP, FM, CS, HB, DN MS, Data Curation – RS; TT, Writing: Original Draft – RS, DN, Writing: Review & Editing – All, Visualization – RS, Project administration & Funding – FM, PP

Acknowledgments

We like to thank the participants. Also, we like to thank the health insurance AOK Nordwest for their help in the certification and implementation of the program, organizing the workshops for coaches, financial support for designing a course booklet and collecting paper-back questionnaires from the course participants. We also like to thank everyone who participated in data collection but did not meet the criteria for authorship.

Literature

[1] Amrhein, V., Trafimow, D., and Greenland, S. 2019. Inferential Statistics as Descriptive Statistics: There Is No Replication Crisis if We Don't Expect Replication. *The American Statistician* 73, sup1, 262–270.

[2] Beck, A. T., Steer, R. A., and Brown, G. 1996. *PsycTESTS Dataset*.

[3] Blosch, C., Schäfer, R., Marées, M. de, and Platen, P. 2019. Comparative analysis of postural control and vertical jump performance between three different measurement devices. *PloS one* 14, 9, e0222502.

[4] Booth, J., Moseley, G. L., Schiltenwolf, M., Cashin, A., Davies, M., and Hübscher, M. 2017. Exercise for chronic musculoskeletal pain: A biopsychosocial approach. *Musculoskeletal care*, 413–421.

[5] Bürkner, P.-C. 2017. brms : An R Package for Bayesian Multilevel Models Using Stan. *J. Stat. Soft.* 80, 1.

[6] Cane, D., Nielson, W. R., and Mazmanian, D. 2018. Patterns of pain-related activity: replicability, treatment-related changes, and relationship to functioning. *Pain* 159, 12, 2522–2529.

[7] Cochrane, A. L. 1972. Effectiveness and efficiency: random reflections on health services.

[8] Dagenais, S., Caro, J., and Haldeman, S. 2008. A systematic review of low back pain cost of illness studies in the United States and internationally. *The spine journal : official journal of the North American Spine Society* 8, 1, 8–20.

[9] Elsey, J. 2021. *Powerful sequential designs using Bayesian estimation: A power analysis tutorial using brms, the tidyverse, and furrr*.

[10] Enrique, B. and Marta, B. 2020. Efficacy, Effectiveness and Efficiency in the Health Care: The Need for an Agreement to Clarify its Meaning. *Int Arch Public Health Community Med* 4, 1.

[11] Fehrmann, E., Tuechler, K., Kienbacher, T., Mair, P., Spreitzer, J., Fischer, L., Kollmitzer, J., and Ebenbichler, G. 2017. Comparisons in Muscle Function and Training Rehabilitation Outcomes Between Avoidance-Endurance Model Subgroups. *The Clinical journal of pain* 33, 10, 912–920.

[12] Foster, N. E., Anema, J. R., Cherkin, D., Chou, R., Cohen, S. P., Gross, D. P., Ferreira, P. H., Fritz, J. M., Koes, B. W., Peul, W., Turner, J. A., Maher, C. G., Buchbinder, R., Hartvigsen, J., Underwood, M., van Tulder, M., Cohen, S. P., Menezes Costa, L., Croft, P., Ferreira, M., Genevay, S., Hancock, M. J., Hoy, D., Karppinen, J., Kongsted, A., Louw, Q., Öberg, B., Peul, W. C., Pransky, G., Schoene, M., Sieper, J., Smeets, R. J., and Woolf, A. 2018. Prevention and treatment of low back pain: evidence, challenges, and promising directions. *The Lancet* 391, 10137, 2368–2383.

[13] Hasenbring, M. 2000. Attentional control of pain and the process of chronification. In *Progress in brain research*. Elsevier, 525–534.

[14] Hasenbring, M. I., Hallner, D., Klasen, B., Streitlein-Böhme, I., Willburger, R., and Rusche, H. 2012. Pain-related avoidance versus endurance in primary care patients with subacute back pain: psychological characteristics and outcome at a 6-month follow-up. *Pain* 153, 1, 211–217.

[15] Hasenbring, M. I., Hallner, D., and Rusu, A. C. 2009. Fear-avoidance- and endurance-related responses to pain: development and validation of the Avoidance-Endurance Questionnaire (AEQ). *European journal of pain (London, England)* 13, 6, 620–628.

[16] Hasenbring, M. I. and Verbunt, J. A. 2010. Fear-avoidance and endurance-related responses to pain: new models of behavior and their consequences for clinical practice. *The Clinical journal of pain* 26, 9, 747–753.

[17] Hides, J. A., Donelson, R., Lee, D., Prather, H., Sahrmann, S. A., and Hodges, P. W. 2019. Convergence and Divergence of Exercise-Based Approaches That Incorporate Motor Control for the Management of Low Back Pain. *The Journal of orthopaedic and sports physical therapy* 49, 6, 437–452.

[18] 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, 6, 453–463.

[19] Kamper, S. J., Apeldoorn, A. T., Chiarotto, A., Smeets, R. J. E. M., Ostelo, R. W. J. G., Guzman, J., and van Tulder, M. W. 2014. Multidisciplinary biopsychosocial rehabilitation for chronic low back pain. *The Cochrane database of systematic reviews,* 9, CD000963.

[20] Korff, M. von, Ormel, J., Keefe, F. J., and Dworkin, S. F. 1992. Grading the severity of chronic pain. *Pain* 50, 2, 133–149.

[21] Kühner, C., Bürger, C., Keller, F., and Hautzinger, M. 2007. Reliabilität und Validität des revidierten Beck-Depressionsinventars (BDI-II). Befunde aus deutschsprachigen Stichproben. *Der Nervenarzt* 78, 6, 651–656.

[22] Liew, B. X. W., Ford, J. J., Scutari, M., and Hahne, A. J. 2021. How does individualised physiotherapy work for people with low back pain? A Bayesian Network analysis using randomised controlled trial data. *PLOS ONE* 16, 10, e0258515.

[23] Luque-Suarez, A., Martinez-Calderon, J., and Falla, D. 2019. Role of kinesiophobia on pain, disability and quality of life in people suffering from chronic musculoskeletal pain: a systematic review. *British journal of sports medicine* 53, 9, 554–559.

[24] Montoya, A. K. and Hayes, A. F. 2017. Two-condition within-participant statistical mediation analysis: A path-analytic framework. *Psychological Methods* 22, 1, 6–27.

[25] Murillo, C., Vo, T.-T., Vansteelandt, S., Harrison, L. E., Cagnie, B., Coppieters, I., Chys, M., Timmers, I., and Meeus, M. 2022. How do psychologically based interventions for chronic musculoskeletal pain work? A systematic review and meta-analysis of specific moderators and mediators of treatment. *Clinical psychology review* 94, 102160.

[26] Niederer, D., Engel, T., Vogt, L., Arampatzis, A., Banzer, W., Beck, H., Moreno Catalá, M., Brenner-Fliesser, M., Güthoff, C., Haag, T., Hönning, A., Pfeifer, A.-C., Platen, P., Schiltenwolf, M., Schneider, C., Trompeter, K., Wippert, P.-M., and Mayer, F. 2020. Motor Control Stabilisation Exercise for Patients with Non-Specific Low Back Pain: A Prospective Meta-Analysis with Multilevel Meta-Regressions on Intervention Effects. *JCM* 9, 9, 3058.

[27] Niederer, D., Vogt, L., Wippert, P.-M., Puschmann, A.-K., Pfeifer, A.-C., Schiltenwolf, M., Banzer, W., and Mayer, F. 2016. Medicine in spine exercise (MiSpEx) for nonspecific low back pain patients: study protocol for a multicentre, single-blind randomized controlled trial. *Trials* 17, 1, 507.

[28] Pietsch, K., Hoyler, A., Frühe, B., Kruse, J., Schulte-Körne, G., and Allgaier, A.-K. 2012. Früherkennung von Depressionen in der Pädiatrie: Kriteriumsvalidität des Beck Depressions-Inventar Revison (BDI-II) und des Beck Depressions-Inventar-Fast Screen for Medical Patients (BDI-FS). *Psychotherapie, Psychosomatik, medizinische Psychologie* 62, 11, 418–424.

[29] R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. https://www.r-project.org/.

[30] Schäfer, R., Schäfer, H., and Platen, P. 2022. Perturbation-based trunk stabilization training in elite rowers: A pilot study. *PLOS ONE* 17, 5, e0268699.

[31] Singal, A. G., Higgins, P. D. R., and Waljee, A. K. 2014. A primer on effectiveness and efficacy trials. *Clinical and translational gastroenterology* 5, e45.

[32] Tagliaferri, S. D., Mitchell, U. H., Saueressig, T., Owen, P. J., Miller, C. T., and Belavy, D. L. 2022. Classification Approaches for Treating Low Back Pain Have Small Effects That Are Not Clinically Meaningful: A Systematic Review With Meta-analysis. *The Journal of orthopaedic and sports physical therapy* 52, 2, 67–84.

[33] Tennant, P. W. G., Arnold, K. F., Ellison, G. T. H., and Gilthorpe, M. S. 2021. Analyses of 'change scores' do not estimate causal effects in observational data. *International journal of epidemiology*.

[34] Thoemmes, F. 2015. Reversing Arrows in Mediation Models Does Not Distinguish Plausible Models. *Basic and Applied Social Psychology* 37, 4, 226–234.

[35] Titze, C., Fett, D., Trompeter, K., Platen, P., Gajsar, H., and Hasenbring, M. I. 2021. Psychosocial subgroups in high-performance athletes with low back pain: eustress-endurance is most frequent, distress-endurance most problematic! *Scandinavian journal of pain* 21, 1, 59–69.

[36] Titze, C., Hasenbring, M. I., Kristensen, L., Bendix, L., and Vaegter, H. B. 2021. Patterns of Approach to Activity in 851 Patients With Severe Chronic Pain: Translation and Preliminary Validation of the 9-item Avoidance-Endurance Fast-Screen (AEFS) Into Danish. *The Clinical journal of pain* 37, 3, 226–236.

[37] Treweek, S. and Zwarenstein, M. 2009. Making trials matter: pragmatic and explanatory trials and the problem of applicability. *Trials* 10, 37.

[38] van Breukelen, G. J. P. 2013. ANCOVA Versus CHANGE From Baseline in Nonrandomized Studies: The Difference. *Multivariate Behavioral Research* 48, 6, 895–922.

[39] Wolff, S. V., Willburger, R., Hallner, D., Rusu, A. C., Rusche, H., Schulte, T., and Hasenbring, M. I. 2018. Avoidance-Endurance Fast-Screen (AE-FS) : Inhalts- und Vorhersagevalidität eines 9-Item-Screeninginstruments für Patienten mit unspezifischen subakuten Rückenschmerzen. *Schmerz (Berlin, Germany)* 32, 4, 283–292.