ISSN: 2641-1911
DOI: 10.33552/ANN.2026.18.000945
Archives in Neurology & Neuroscience
Iris Publishers

Research Article

Rethinking Severity-Based Treatment Allocation in Persistent Spinal Pain: Substantial Benefit of Outpatient Interdisciplinary Multimodal Therapy in Highly Burdened Patients

Michael A. Überall[1*], Thomas Nolte[2], Gerhard H.H. Müller-Schwefe[3] and Harry Kletzko[4]

[1]*Private Institute of Neurological Sciences – IFNAP, Center of Excellence in Health Care Research, Nürnberg, Germany*

[2]*Center for Outpatient Palliative Care, Wiesbaden, Germany*

[3]*Center for Interdisciplinary Pain and Palliative Care, Göppingen, Germany*

[4]*Integrative Managed Care GmbH, Limburg, Germany*

***Corresponding author:** Michael A. Überall, Private Institute of Neurological Sciences – IFNAP, Center of Excellence in Health Care Research, Nordostpark 51, 90411 Nürnberg, Germany.

Received Date: May 23, 2026

Published Date: June 16, 2026

Abstract

Background: Interdisciplinary multimodal pain therapy is a recommended treatment approach for persistent spinal pain; however, it remains unclear whether treatment effectiveness differs according to baseline severity, particularly in outpatient settings. Patients with high biopsychosocial burden are frequently considered candidates for more intensive treatment pathways, although evidence supporting severity-based allocation remains limited.

Methods: This retrospective multicenter cohort study analyzed 9,217 patients with persistent spinal pain participating in a structured three-week interdisciplinary outpatient multimodal pain program following non-confirmation of a surgical indication. Baseline severity was assessed using a multidimensional biopsychosocial composite score (0–40 points) and categorized back pain severity into five severity groups (BPSG 1–5). Outcomes across multiple domains were harmonized and standardized using baseline-derived harmonized standard deviation scores (hSDS). Treatment response was evaluated using standardized change scores (Δ-hSDS), analysis of covariance models with baseline adjustment, and responder analyses across predefined thresholds of improvement.

Results: Baseline severity demonstrated a clear gradient across groups. Standardized change scores showed a strong inverse association with baseline values (R^2 = 0.56), indicating larger improvements in patients with greater baseline impairment. Median Δ-hSDS increased progressively across severity groups, ranging from approximately 0 in the lowest severity group to >1.5 in the highest severity group. In baseline-adjusted models, severity group remained significantly associated with Δ-hSDS (partial η^2 =0.058), whereas baseline values showed a smaller independent effect. Responder analyses confirmed increasing proportions of patients achieving large standardized improvements across higher severity groups. Although absolute week-3 outcomes remained associated with baseline status, substantial short-term improvements were observed even in patients with severe and extreme biopsychosocial burden.

Conclusions: Patients with high and very high baseline impairment demonstrated substantial short-term benefit from structured interdisciplinary outpatient multimodal pain therapy. High biopsychosocial burden should therefore not be interpreted as evidence against outpatient multimodal treatment or as an automatic indication for inpatient care. These findings support outpatient interdisciplinary treatment as a relevant component of stepped-care approaches across a broad range of baseline severity.

Keywords: Persistent spinal pain; Multimodal pain therapy; Outpatient treatment; Interdisciplinary care; Biopsychosocial model; Severity stratification; Harmonized standard deviation scores; Treatment response; Stepped care; chronic pain

Introduction

Persistent spinal pain is among the leading causes of long-term disability worldwide and represents a substantial burden for patients, healthcare systems, and society [1,2]. Despite considerable advances in diagnostic imaging and surgical techniques, the relationship between structural findings and clinical symptoms remains inconsistent [3,4]. A substantial proportion of patients referred for spinal surgery present with persistent pain syndromes that cannot be adequately explained by identifiable structural pathology alone. This discrepancy frequently complicates treatment decisions and contributes to considerable variability in clinical management.

In recent years, interdisciplinary multimodal pain therapy has become an established component in the management of persistent spinal pain. By integrating medical, physiotherapeutic, psychological, and educational interventions, multimodal approaches aim to address the complex biological, psychological, and social factors contributing to pain chronification. Current national and international guidelines recommend interdisciplinary multimodal treatment particularly for patients with persistent, therapy-resistant, and functionally disabling pain conditions [5-7].

However, an important unresolved question concerns the relationship between baseline severity and treatment effectiveness. In routine care, patients with pronounced biopsychosocial burden are often assumed to require more intensive treatment settings, including inpatient multimodal programs. High levels of disability, chronicity, psychological distress, and social impairment are frequently regarded as indicators for escalation of care [8-11]. Consequently, patients with severe or extreme baseline impairment may be considered less suitable for outpatient treatment pathways. Despite the clinical relevance of this assumption, empirical evidence supporting severity-based allocation decisions remains limited.

This issue is particularly relevant in patients undergoing second-opinion assessment prior to planned spinal surgery. In many cases, surgical indication is ultimately not confirmed, creating a need for effective conservative treatment strategies [12-15]. These patients frequently present with complex clinical profiles characterized by substantial functional limitations, maladaptive pain-related beliefs, psychological distress, and long-standing pain histories. Determining whether structured outpatient interdisciplinary treatment remains effective across varying levels of baseline burden is therefore highly relevant for treatment planning and resource allocation.

A further challenge in evaluating treatment effectiveness across heterogeneous patient populations lies in the operationalization of baseline severity itself. Previous studies have often relied on single- domain measures such as pain intensity, disability indices, or classifications such as the Chronic Pain Grading Scale [16-18]. While these instruments provide valuable information, they do not fully reflect the multidimensional nature of persistent spinal pain. Furthermore, comparisons between severity groups are methodologically complicated by differences in scale properties, ceiling effects, and regression-to-the-mean phenomena [19,20].

To address these limitations, multidimensional severity models may provide a more comprehensive framework by integrating biological, functional, psychological, and social dimensions of chronic pain [6,9,10,12,17]. Likewise, harmonized standardization procedures allow outcomes derived from instruments with different scaling properties to be expressed on a common metric and facilitate cross-domain comparisons [21-23].

Objectives

The present study therefore aimed to investigate the relationship between multidimensional biopsychosocial baseline severity and treatment-associated changes in a large multicenter cohort of patients with persistent spinal pain participating in a structured interdisciplinary outpatient multimodal pain program after non-confirmation of surgical indication. A rule-based biopsychosocial severity score was developed to classify patients into five ordered severity groups. Treatment response was assessed using harmonized standardized outcome measures, baseline-adjusted analyses, and responder approaches. We hypothesized that treatment-associated improvements would occur across all severity levels and specifically examined whether patients with high and very high baseline burden demonstrate clinically meaningful benefit within an outpatient interdisciplinary treatment setting.

Methods

Study design and setting

This study represents a retrospective cohort analysis of prospectively collected routine care data derived from a nationwide interdisciplinary integrated care network in Germany. The network provides structured assessment and treatment pathways for patients with persistent spinal pain and includes multiple specialized centers applying standardized diagnostic and therapeutic procedures.[12].

The study period extended from January 2014 to December 2025. All patients underwent interdisciplinary second-opinion assessment prior to planned spinal surgery. Evaluations were performed by experienced interdisciplinary pain specialist teams and included medical, functional, and psychosocial assessment procedures. Patients in whom surgical indication was not confirmed and who subsequently participated in a structured interdisciplinary outpatient multimodal pain program were eligible for inclusion.

Participants

The study population consisted of all adult patients with persistent spinal pain who completed both baseline assessment before initiation of treatment and follow-up assessment at the end of the three- week outpatient program.

Patients with incomplete information required for severity score calculation or harmonized outcome score generation were excluded from analysis.

A total of 9,217 patients fulfilled the inclusion criteria and were available for analysis.

Intervention

The intervention consisted of a structured interdisciplinary outpatient multimodal pain therapy program delivered over three weeks within a standardized integrated care framework. The treatment program was embedded within a nationwide interdisciplinary care network following an ambulatory-first treatment philosophy.

Prior to treatment initiation, all patients underwent a standardized interdisciplinary bio-psycho-social assessment including medical, functional, and psychological evaluation procedures. Treatment recommendations were subsequently determined through structured interdisciplinary case conferences and shared decision-making processes.

Rather than representing a rigid one-size-fits-all intervention, the treatment framework followed an adaptive architecture in which therapeutic components and treatment intensity were aligned with individual clinical complexity and biopsychosocial burden.

The program integrated coordinated medical, physiotherapeutic, psychological, and educational treatment components and was delivered by interdisciplinary teams including physicians, psychologists, physiotherapists, and additional healthcare professionals where required. Treatment focused on functional restoration, improvement of pain-related coping, modification of maladaptive beliefs and behaviors, promotion of self-management strategies, and reduction of pain- related disability.

To ensure comparability across participating centers, treatment structure and core components followed predefined standards within the integrated care network and were continuously coordinated across participating sites.

Development of the biopsychosocial severity groups (BPSG)

Baseline severity was operationalized using a multidimensional biopsychosocial severity model integrating biological, functional, psychological, and social domains of persistent pain.

Twenty baseline variables were included: pain intensity index, Chronic Pain Grading Scale (CPGS), modified Pain Disability Index (mPDI), Hannover Functional Ability Questionnaire (FFbHR), VR-12 physical and mental component scores (PCS and MCS), Mainzer Fragebogen zur Handfunktion (MFHW), Depression Anxiety Stress Scales (DASS-21) subscales for depression, anxiety, and stress, Fear-Avoidance Beliefs Questionnaire (FABQ) work and physical activity subscales, pain self-efficacy (FESS/PSEQ), Mainz Pain Staging System (MPSS), painDETECT questionnaire (PDQ-7), pain duration, AVEM classification, prior surgery status, degree of disability (GdB), and work incapacity status.

Each variable was transformed a priori into an ordinal severity score of 0, 1, or 2 points representing low, moderate, or high impairment according to predefined thresholds derived from established cut-offs or clinically meaningful ranges (see Figure 1). Categorization into three ordinal levels was chosen to preserve interpretability and maintain comparable weighting across heterogeneous domains while reducing the influence of individual scale-specific distributions.

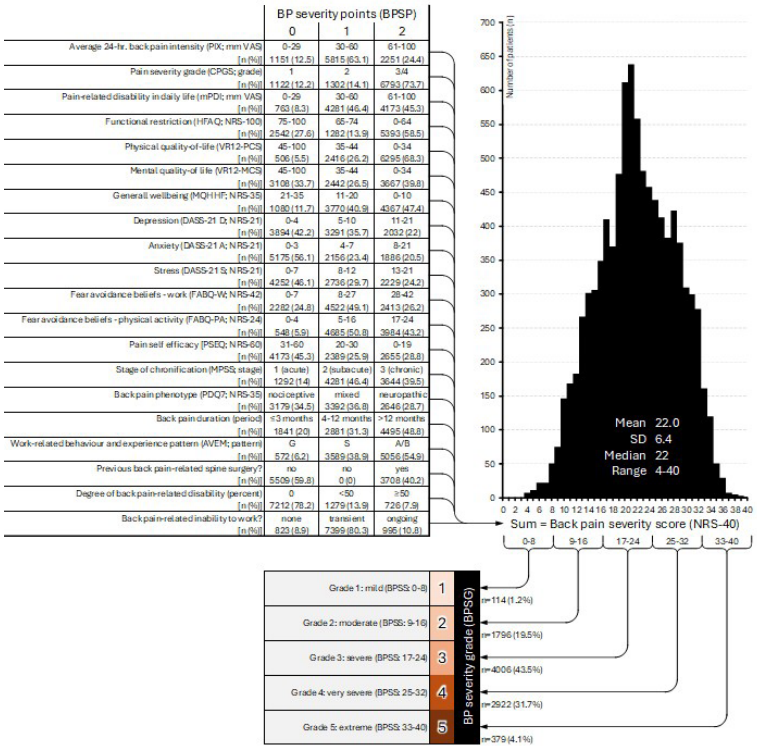

Figure 1: Development and distribution of the biopsychosocial severity groups (BPSG).
The multidimensional biopsychosocial severity model was derived from predefined clinical domains including pain intensity, pain-related disability, functional impairment, psychological distress, fear-avoidance beliefs, self-efficacy, pain chronicity, and social-medical indicators. Individual variables were transformed into ordinal scores (0–2 points) and aggregated into a composite severity score ranging from 0 to 40 points. Patients were subsequently classified into five biopsychosocial severity groups (BPSG 1–5), representing increasing levels of multidimensional baseline burden.

To ensure directional consistency across all domains, scoring was harmonized so that increasing values consistently represented increasing biopsychosocial burden. Individual point values were summed to obtain a composite score ranging from 0 to 40. Patients were classified into five predefined biopsychosocial back pain severity groups: BPSG 1: 0–8 points, BPSG 2: 9–16 points, BPSG 3: 17–24 points, BPSG 4: 25–32 points, and BPSG 5: 33–40 points, representing increasing levels of multidimensional baseline burden.

Harmonized standardized outcome measures (hSDS)

Because included instruments differed substantially regarding scale range, orientation, and measurement properties, harmonized standard deviation scores (hSDS) were used to allow cross- domain comparison. For each instrument, baseline mean values and standard deviations of the total cohort were used to derive standardized scores at baseline (BL-hSDS) and after completion of treatment (W3-hSDS). Before standardization, all variables were harmonized to ensure that higher values consistently reflected more favorable clinical status. Composite hSDS values were calculated by averaging standardized values across all included domains.

Treatment-associated change was quantified as: Δ-hSDS = W3-hSDS − BL-hSDS. Positive values represented improvement, whereas negative values represented deterioration.

Statistical analysis

Descriptive statistics were calculated for baseline characteristics and outcome variables across severity groups. Relationships between BL-hSDS and treatment-associated change (Δ-hSDS) were examined using linear regression analyses and graphical inspection. Differences in change scores across severity groups were evaluated using analysis of variance and visualized using boxplots. Responder analyses categorized patients according to predefined thresholds of standardized improvement expressed in multiples of standard deviations.

To account for baseline differences, analyses of covariance (ANCOVA) were performed with two

separately specified models:

1. Δ-hSDS as dependent variable with severity group, BL-hSDS, and BL-hSDS × severity interaction included as predictors.

2. W3-hSDS as dependent variable with severity group and BL-hSDS included as predictors.

Estimated marginal means (EMMs) adjusted for baseline values were calculated for both models. Effect sizes were reported as partial eta-squared (η^2). To account for the large number of secondary outcomes, adjustment for multiple testing was performed using the Bonferroni method. All statistical tests were two-sided. A p-value <0.05 was considered statistically significant for the primary endpoint, whereas Bonferroni-adjusted thresholds were applied for secondary analyses.

All statistical analyses were performed using SPSS (PASW Statistics, Version 18.0; SPSS Inc., Chicago,

IL, USA).

Ethical and Regulatory Considerations

The integrated care program evaluated in the present study was conducted within the framework of contractual integrated care agreements according to §140a SGB V (German Social Code, Book V). Participation in the program was voluntary and required written informed consent prior to initiation of the interdisciplinary second-opinion assessment. Patients provided consent both for participation in the structured care program and for the use of pseudonymized clinical data for healthcare research purposes. This consent was documented in written form and subsequently confirmed electronically at first use of the German Pain e-Registry (GPeR).

Clinical data were stored in pseudonymized form within the registry infrastructure. For the present analysis, fully anonymized datasets were extracted. The investigation represents a retrospective analysis of routinely collected healthcare data and did not involve any additional diagnostic or therapeutic interventions beyond those applied as part of the specialized multimodal care program.

Ethical review and approval were waived because the study represents a secondary analysis of anonymized routine care data collected within a structured integrated care program according to §140a SGB V. Nevertheless, the protection of patient rights and data privacy in the context of the present analysis was reviewed and monitored by the Ethics Board of the German Pain League (Deutsche Schmerzliga).

The study was conducted in accordance with the ethical principles outlined in the Declaration of Helsinki.

Funding and Governance

The present study received no external public or commercial research funding. The analysis was conducted on behalf of Integrative Managed Care GmbH (IMC GmbH) within the framework of its structured healthcare evaluation activities.

Statistical analyses were performed independently by O.Meany-MDPM GmbH under methodological supervision of the Institute for Neurological Sciences (IFNAP) and the Center of Excellence for Health Care Research, Nürnberg, Germany.

The statistical analysis plan was defined prior to data evaluation. All analyses were conducted independently of the network operator. Interpretation of results and manuscript preparation were performed by the authors, all of whom had full access to the anonymized dataset and critically reviewed and approved the final manuscript.

The project was prospectively registered within the European Medicines Agency (EMA) framework as a healthcare research initiative and documented within the European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP) registry to ensure transparency of study objectives and methodology (EU PAS identifier: EUPAS1000000952).

Results

Study population and biopsychosocial severity stratification

A total of 9,217 patients met the inclusion criteria and were available for analysis. Based on the multidimensional biopsychosocial severity model, patients were classified into five predefined biopsychosocial severity groups (BPSG), representing increasing levels of baseline burden. Group sizes were distributed as follows: BPSG 1 (n = 114), BPSG 2 (n = 1,796), BPSG 3 (n = 4,006), BPSG 4 (n =2,922), and BPSG 5 (n = 379).

Baseline demographic and clinical characteristics stratified by severity group are presented in Table 1. The distribution of patients across groups demonstrated the expected predominance of intermediate severity categories, whereas the lowest and highest categories comprised smaller patient numbers (Table 1).

Table 1: Baseline characteristics of the study population stratified by biopsychosocial severity group.

Cohort Patients [n (%)]	All 9217 (100.0)	(Grade 1) 114 (1.2)	(Grade 2) 1796 (19.5)	(Grade 3) 4006 (43.5)	(Grade 4) 2922 (31.7)	(Grade 5) 379 (4.1)	Between cohorts significance (eta2)
Average 24-hr. pain intensity [PIX; mm VAS; mean (SD)]	49.5 (15.8)	32.5 (15.1)	37.2 (14.3)	48.4 (14.0)	57.4 (13.1)	64.6 (11.9)	<0.001 (0.252)
Modified Pain Disability Index [mPDI; mm VAS; mean (SD)]	56.7 (18.0)	27.6 (10.3)	39.7 (14.5)	54.9 (14.6)	68.4 (13.7)	75.0 (11.1)	<0.001 (0.386)
Hannover Functional Assessment Questionnaire [HFAQ; NRS-100; mean (SD)]	57.6 (23.8)	88.1 (15.8)	77.5 (17.6)	59.5 (21.3)	44.1 (20.0)	36.4 (16.9)	<0.001 (0.294)
Physical quality-of-life [VR12-PCS; mean (SD)]	30.4 (9.3)	45.0 (5.8)	37.2 (8.7)	29.9 (8.6)	27.1 (7.9)	24.4 (6.9)	<0.001 (0.195)
Mental quality-of-life [VR12-MCS; mean (SD)]	39.5 (11.8)	56.4 (5.3)	47.9 (10.3)	41.8 (10.5)	32.1 (8.8)	27.8 (7.6)	<0.001 (0.306)
Generall wellbeing [MQHHF; NRS-35; mean (SD)]	11.6 (7.2)	25.4 (5.7)	18.0 (6.8)	11.9 (6.0)	7.4 (5.1)	5.1 (3.8)	<0.001 (0.341)
Depression [DASS-21 D; NRS-21; mean (SD)]	6.6 (5.2)	0.7 (1.1)	2.7 (2.7)	5.1 (3.7)	10.4 (5.0)	13.3 (4.2)	<0.001 (0.401)
Anxiety [DASS-21 A; NRS-21; mean (SD)]	4.3 (4.7)	0.8 (1.1)	1.5 (2.0)	2.7 (2.8)	7.5 (5.5)	10.9 (4.0)	<0.001 (0.347)
Stress [DASS-21S; NRS-21; mean (SD)]	8.6 (5.3)	2.2 (2.1)	4.7 (3.6)	7.1 (4.0)	12.7 (4.5)	14.0 (3.6)	<0.001 (0.398)
Fear avoidance beliefs work [FABQ-W; NRS-42; mean (SD)]	18.1 (12.2)	5.3 (6.1)	10.1 (8.8)	16.3 (11.4)	24.7 (11.2)	28.4(9.3)	<0.001 (0.229)
Fear avoidance beliefs - physical activity (FABQ-PA; NRS-24; mean (SD)]	14.9 (5.9)	7.6 (5.0)	11.8 (5.5)	14.5 (5.7)	17.2 (5.1)	18.3 (4.9)	<0.001 (0.138)
Pain self efficacy [PSEQ; NRS-60; mean (SD)]	29.6 (13.3)	52.2 (10.2)	39.1 (13.5)	30.0 (12.3)	23.9 (10.3)	17.7 (5.2)	<0.001 (0.228)
Pain phenotype [PDQ7; NRS-35; mean (SD)]	13.9 (7.3)	6.9 (4.4)	8.9 (5.8)	12.7 (6.4)	17.9 (6.8)	20.7 (5.7)	<0.001 (0.246)
Harmonized standard deviation scores (hSDS): mean	0	1.302	0.736	0.137	-0.568	-0.951	<0.001 (0.745)
standard deviation	0.606	0.264	0.263	0.274	0.367	0.302	
median	0.023	1.327	0.73	0.129	-0.527	-0.96	
minimum	-1.768	0.738	-0.098	-0.731	-1.66	-1.768	
maximum	1.772	1.772	1.554	1.079	0.331	-0.282	
95% confidence interval	0.014	0.055	0.014	0.01	0.015	0.035	

Descriptive demographic, clinical, functional, psychological, and social-medical characteristics at baseline across the five predefined biopsychosocial severity groups (BPSG 1–5). Higher severity categories represent increasing multidimensional biopsychosocial burden. Group comparisons are presented with corresponding significance levels and effect sizes.

Baseline characteristics demonstrated a progressive gradient across severity groups. Patients in higher BPSG categories exhibited increasing impairment across multiple biopsychosocial domains, including pain-related disability, psychological distress, functional limitation, fear-avoidance beliefs, pain chronicity, and social-medical indicators. The derivation and operationalization of the multidimensional severity score are summarized in Table 2.

Table 2: Week-3 characteristics of the study population stratified by biopsychosocial severity group (BPSG).

Cohort Patients [n (%)]	All 9217 (100.0)	(Grade 1) 114 (1.2)	(Grade 2) 1796 (19.5)	(Grade 3) 4006 (43.5)	(Grade 4) 2922 (31.7)	(Grade 5) 379 (4.1)	Between cohorts significance (eta2)
Average 24-hr. pain intensity [PIX; mm VAS; mean (SD)]	23.1 (17.3)	22.1 (15.9)	21.1 (15.9)	22.6 (17.0)	24.5 (18.1)	26.2 (19.8)	<0.001 (0.006)
Modified Pain Disability Index [mPDI; mm VAS; mean (SD)]	26.6 (20.5)	23.0 (20.4)	23.6 (19.6)	26.3 (19.9)	28.7 (21.3)	30.2 (22.3)	<0.001 (0.009)
Hannover Functional Assessment Questionnaire [HFAQ; NRS-100; mean (SD)]	74.3 (19.4)	89.3 (9.6)	82.7 (14.0)	75.1 (18.3)	69.4 (20.3)	59.7 (26.6)	<0.001 (0.088)
Physical quality-of-life [VR12-PCS; mean (SD)]	43.7 (10.0)	44.8 (10.1)	44.4 (9.4)	43.6 (10.1)	43.4 (10.1)	43.1 (10.6)	<0.001 (0.001)
Mental quality-of-life [VR12-MCS; mean (SD)]	46.7 (10.4)	47.8 (9.6)	48.1 (10.2)	46.8 (10.2)	46.0 (10.5)	44.7 (10.9)	<0.001 (0.007)
Generall wellbeing [MQHHF; NRS-35; mean (SD)]	19.5 (7.8)	27.8 (6.3)	23.9 (6.5)	20.0 (6.9)	16.7 (7.8)	12.3 (7.3)	<0.001 (0.154)
Depression [DASS-21 D; NRS-21; mean (SD)]	3.3 (3.3)	2.7 (2.9)	3.0 (3.1)	3.2 (3.2)	3.6 (3.5)	4.1 (3.9)	<0.001 (0.007)
Anxiety [DASS-21 A; NRS-21; mean (SD)]	2.0 (2.1)	1.6 (1.7)	1.8 (1.9)	1.9 (1.9)	2.2 (2.2)	2.8 (2.9)	<0.001 (0.011)
Stress [DASS-21S NRS-21; mean (SD)]	3.9 (3.5)	3.7 (3.3)	3.5 (3.2)	3.8 (3.4)	4.2 (3.7)	4.7 (4.3)	<0.001 (0.007)
Fear avoidance beliefs - work [FABQ-W; NRS-42; mean (SD)]	12.9 (9.8)	7.3 (7.2)	10.2 (7.9)	11.7 (9.6)	15.8 (9.9)	18.6 (11.3)	<0.001 (0.068)
Fear avoidance beliefs - physical activity (FABQ-PA; NRS-24; mean (SD)]	9.7 (6.0)	6.1 (5.1)	8.4 (5.7)	9.6 (6.1)	10.4 (5.8)	11.7 (6.6)	<0.001 (0.022)
Pain self efficacy [PSEQ; NRS-60; mean (SD)]	46.0 (10.3)	46.7 (10.2)	46.8 (10.0)	46.0 (10.1)	45.8 (10.6)	43.8 (11.7)	<0.001 (0.003)
Pain phenotype [PDQ7; NRS-35; mean (SD)]	9.9 (6.0)	6.0 (3.9)	7.9 (5.2)	9.3 (5.5)	11.6 (6.3)	13.1 (7.4)	<0.001 (0.067)
Harmonized standard deviation scores (hSDS): mean	0.945	1.266	1.127	0.976	0.819	0.627	<0.001 (0.09)
standard deviation	0.439	0.4	0.36	0.408	0.446	0.559	
median	0.989	1.306	1.151	1.004	0.874	0.685	
minimum	-1.671	0.152	-0.421	-0.675	-1.115	-1.671	
maximum	2.105	2.079	2.038	2.105	1.919	1.696	
95% confidence interval	0.009	0.084	0.019	0.014	0.018	0.064	

Descriptive outcome measures obtained after completion of the three-week interdisciplinary 749 outpatient multimodal pain program across the five predefined biopsychosocial severity groups. 750 Harmonized standard deviation scores (hSDS) are included as multidimensional standardized 751 outcome measures.

Baseline standardized status and treatment-associated change

Baseline harmonized standardized scores (BL-hSDS), week-3 harmonized standardized scores (W3-hSDS), and corresponding treatment-associated changes (Δ-hSDS) are summarized in Table 3.

BL-hSDS values demonstrated a clear distribution across sever-

ity categories, confirming the expected ordering of the severity classification. Lower severity groups showed more favorable baseline standardized scores, whereas patients assigned to higher severity categories demonstrated progressively lower BL-hSDS values.

Table 3: Treatment-associated changes in harmonized standard deviation scores (Δ-hSDS) according to biopsychosocial severity group (BPSG).

Cohort Patients [n (%)]	All 9217 (100.0)	Back pain severity grade (BPSG)					Between cohorts significance (eta2)
		Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	
		114 (1.2)	1796 (19.5)	4006 (43.5)	2922 (31.7)	379 (4.1)	
Average 24-hr. pain intensity [PIX; mm VAS; mean (SD)]	26.5 (21.9)	10.5 (22.0)	16.0 (20.5)	25.8 (20.8)	32.9 (21.1)	38.4 (22.7)	<0.001 (0.091)
Modified Pain Disability Index [mPDI; mm VAS; mean (SD)]	30.1 (26.0)	4.5 (21.9)	16.1 (24.1)	28.6 (23.7)	39.7 (25.1)	44.8 (24.7)	<0.001 (0.127)
Hannover Functional Assessment Questionnaire [HFAQ; NRS-100; mean (SD)]	16.7 (23.9)	1.2 (18.7)	5.2 (19.3)	15.6 (22.6)	25.2 (24.2)	23.3 (27.8)	<0.001 (0.095)
Physical quality-of-life [VR12-PCS; mean (SD)]	13.3 (13.4)	-0.2 (11.3)	7.2 (12.7)	13.7 (13.1)	16.4 (12.8)	18.8 (12.3)	<0.001 (0.077)
Mental quality-of-life [VR12-MCS; mean (SD)]	7.2 (15.1)	-8.6 (10.9)	0.2 (14.3)	4.9 (14.1)	13.9 (13.6)	16.9 (13.6)	<0.001 (0.144)
General wellbeing [MQHHF; NRS-35; mean (SD)]	7.9 (8.0)	2.4 (5.9)	5.9 (7.5)	8.1 (7.8)	9.3 (8.3)	7.1 (8.1)	<0.001 (0.028)
Depression [DASS-21 D; NRS-21; mean (SD)]	3.3 (5.9)	-2.0 (3.1)	-0.3 (4.1)	1.8 (4.7)	6.8 (6.0)	9.2 (5.6)	<0.001 (0.265)
Anxiety [DASS-21 A; NRS-21; mean (SD)]	2.3 (5.0)	-0.8 (2.0)	-0.3 (2.7)	0.8 (3.3)	5.3 (5.9)	8.1 (4.8)	<0.001 (0.269)
Stress [DASS-21 S; NRS-21; mean (SD)]	4.7 (6.1)	-1.4 (3.9)	1.1 (4.7)	3.3 (5.1)	8.5 (5.8)	9.4 (5.5)	<0.001 (0.251)
Fear avoidance beliefs - work [FABQ-W; NRS-42; mean (SD)]	5.2 (12.2)	-2.0 (4.7)	-0.1 (9.8)	4.6 (11.7)	8.8 (12.9)	9.8 (12.8)	<0.001 (0.076)
Fear avoidance beliefs - physical activity [FABQ-PA; NRS-24; mean (SD)]	5.2 (7.5)	1.5 (4.5)	3.3 (6.9)	4.9 (7.7)	6.8 (7.3)	6.6 (7.8)	<0.001 (0.031)
Pain self efficacy [PSEQ; NRS-60; mean (SD)]	16.4 (16.3)	-5.5 (14.7)	7.6 (16.4)	15.9 (15.4)	21.9 (14.6)	26.0 (12.9)	<0.001 (0.13)
Pain phenotype [PDQ7; NRS-35; mean (SD)]	4.0 (7.8)	0.8 (5.7)	1.0 (7.0)	3.4 (7.2)	6.4 (8.2)	7.6 (8.6)	<0.001 (0.072)
Harmonized standard deviation scores (hSDS): mean	0.945	-0.037	0.391	0.839	1.387	1.578	<0.001 (0.386)
standard deviation	0.633	0.415	0.416	0.453	0.576	0.624	
median	0.921	-0.048	0.407	0.855	1.383	1.603	
minimum	-1.309	-1.118	-1.309	-0.98	-0.564	-0.426	
maximum	3.099	0.853	1.616	2.334	3.099	2.934	
95% confidence interval	0.015	0.087	0.022	0.016	0.024	0.072	

Descriptive statistics of treatment-associated changes expressed as harmonized standard deviation 758 scores (Δ-hSDS) across the five predefined biopsychosocial severity groups (BPSG 1–5). Positive values 759 indicate improvement following completion of the three-week interdisciplinary outpatient 760 multimodal pain program. Mean values, standard deviations, medians, ranges, and confidence 761 intervals are presented.

Following completion of the three-week interdisciplinary outpatient treatment program, W3-hSDS values improved across all groups. However, the magnitude of treatment-associated change differed substantially according to baseline severity.

Standardized change scores (Δ-hSDS) increased progressively across severity categories. Median Δ-hSDS values ranged from approximately no change in the lowest severity category to marked positive changes in the highest categories. The largest standardized improvements were observed in patients classified as BPSG 4 and BPSG 5.

Relationship between baseline status and treatment-associated change

The relationship between baseline status and treatment-associated change is illustrated in Figure 2A. Across the total cohort, BL-hSDS values demonstrated a pronounced inverse association with Δ-hSDS. Patients with lower baseline values generally exhibited larger positive change scores, whereas patients with more favorable baseline values showed smaller changes.

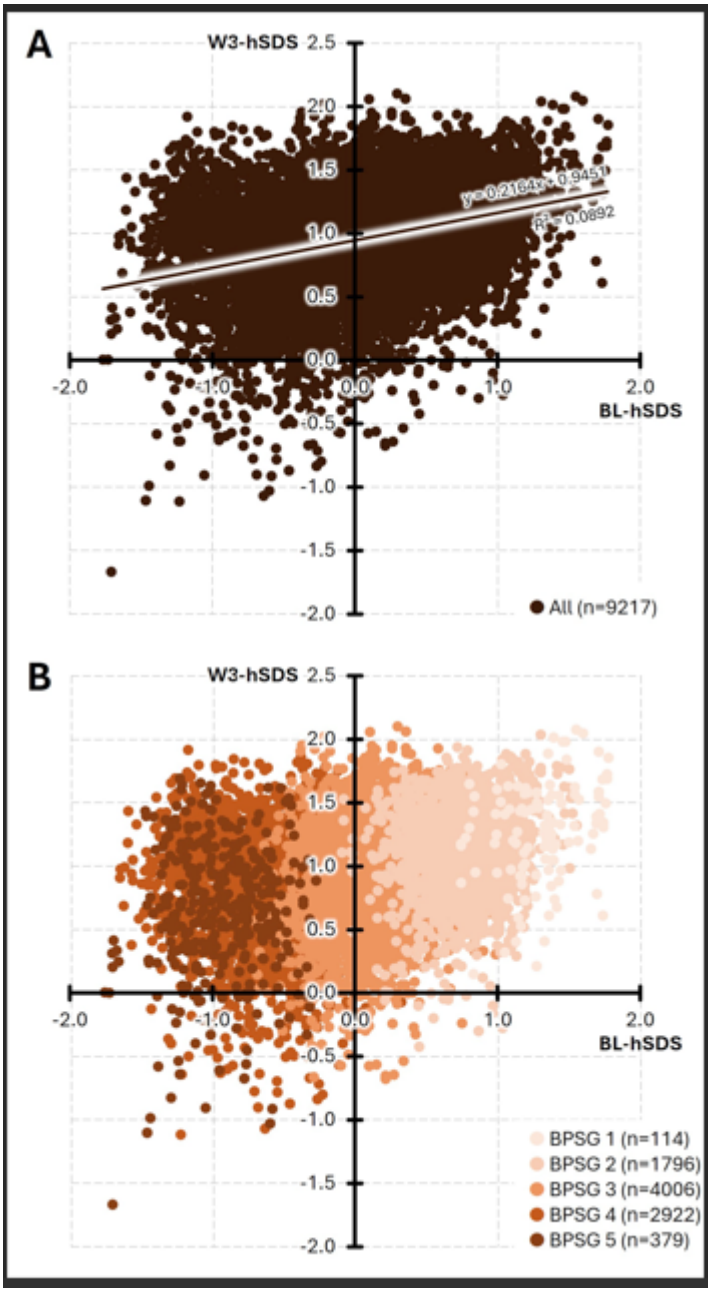

Figure 2: Relationship between baseline and postreatment harmonized standard deviation scores.
(A) Scaterplot illustrating the association between baseline harmonized standard deviation scores (BL-hSDS) and week-3 harmonized standard deviation scores (W3-hSDS) for the total cohort (n = 9,217). The regression line illustrates the relationship between baseline status and absolute postreatment outcome levels.
(B) Scaterplot stratified according to biopsychosocial severity groups (BPSG 1–5), illustrating the distribution of patients across baseline and week-3 standardized outcome values.

Linear regression analysis demonstrated a substantial proportion of explained variance (R^2 = 0.56), indicating a strong relationship between baseline status and standardized change.

When stratified according to biopsychosocial severity groups (Figure 2B), patients were distributed across distinct baseline ranges corresponding to the predefined severity categories. The observed pattern of treatment-associated change followed this distribution, with progressively larger Δ-hSDS values observed in groups with higher baseline impairment.

Distribution of standardized treatment responses

The distribution of treatment-associated change across severity groups is shown in Figure 3A. Median Δ-hSDS values increased progressively from BPSG 1 to BPSG 5. Patients in lower severity

groups demonstrated comparatively small standardized changes and a broader distribution around toward larger positive Δ-hSDS values.

In addition, variability of treatment response increased across higher severity groups, reflected by wider interquartile ranges and broader overall score distributions. Responder analyses further characterized the magnitude of improvement across severity groups (Figure 3B, Table 4). Lower severity groups contained larger proportions of patients showing no change or only small improvements, whereas higher severity groups demonstrated progressively larger proportions of patients achieving substantial standardized improvements.

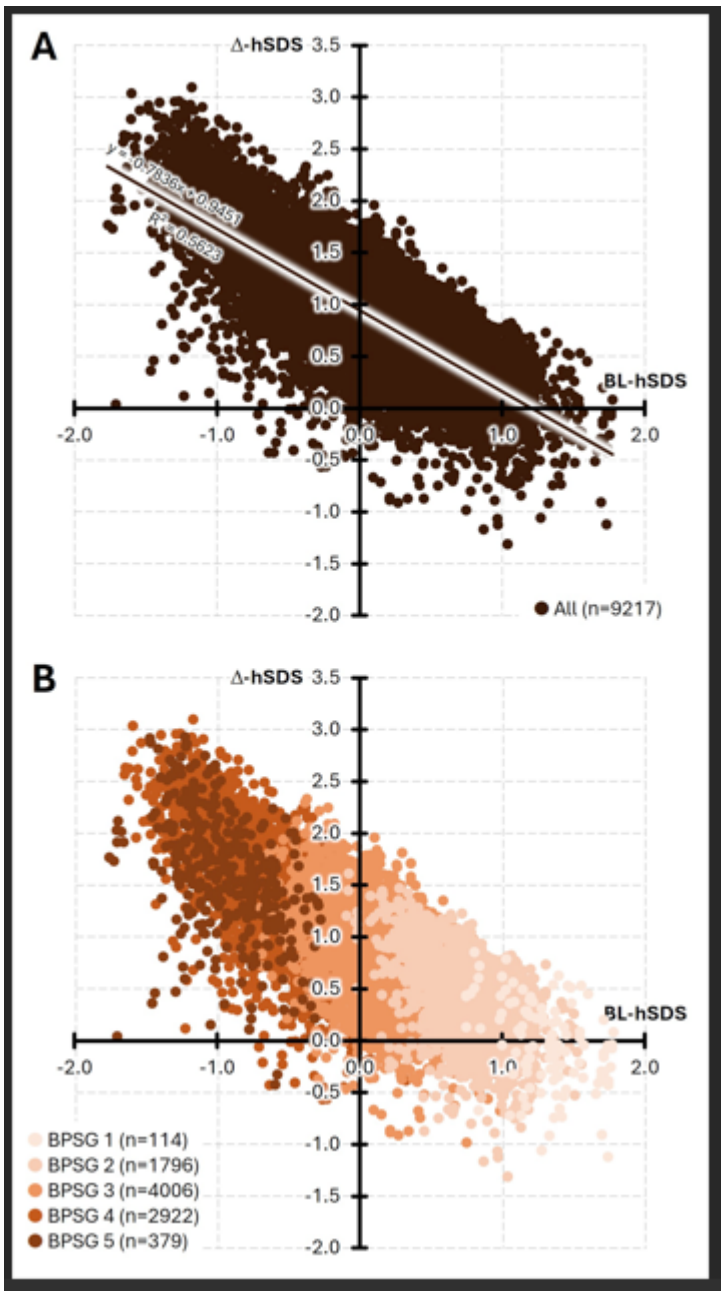

Figure 3: Relationship between baseline status and treatment-associated change.
(A) Scaterplot illustrating the association between baseline harmonized standard deviation scores (BL-hSDS) and treatment-associated changes expressed as Δ-hSDS for the total cohort (n = 9,217). The regression line demonstrates the inverse relationship between baseline status and magnitude of standardized change.
(B) Scaterplot stratified according to biopsychosocial severity groups (BPSG 1–5), illustrating severity-specific distributions of baseline values and corresponding treatment-associated changes.

Table 4: Within-group effect sizes for treatment-associated changes across biopsychosocial severity groups 766 (BPSG).

Cohort Patients [n (%)]	All 9217 (100.0)	Back pain severity grade (BPSG)				
		Grade 1	Grade 2	Grade 3	Grade 4	Grade 5
		114 (1.2)	1796 (19.5)	4006 (43.5)	2922 (31.7)	379 (4.1)
Average 24-hr. pain intensity [PIX; mm VAS; mean (SD)]	1.594	0.671	1.065	1.657	2.082	2.351
Modified Pain Disability Index [mPDI; mm VAS; mean (SD)]	1.379	0.285	0.934	1.639	2.217	2.543
Hannover Functional Assessment Questionnaire [HFAQ; NRS-100; mean (SD)]	0.769	0.092	0.407	0.786	1.256	1.046
Physical quality-of-life [VR12-PCS; mean (SD)]	1.377	0.024	0.795	1.461	1.798	2.091
Mental quality-of-life [VR12-MCS; mean (SD)]	0.647	1.109	0.784	0.483	1.435	1.799
General wellbeing [MQHHF; NRS-35; mean (SD)]	1.052	0.4	0.887	1.253	1.411	1.237
Depression [DASS-21 D; NRS-21; mean (SD)]	0.758	0.912	0.103	0.549	1.576	2.27
Anxiety [DASS-21 A; NRS-21; mean (SD)]	0.632	0.559	0.154	0.334	1.265	2.319
Stress [DASS-21 S; NRS-21; mean (SD)]	1.047	0.542	0.352	0.889	2.063	2.345
Fear avoidance beliefs - work [FABQ-W; NRS-42; mean (SD)]	0.47	0.3	0.012	0.436	0.842	1.476
Fear avoidance beliefs - physical activity (FABQ-PA; NRS-24; mean (SD)]	0.874	0.297	0.607	0.83	1.245	1.135
Pain self efficacy [PSEQ; NRS-60; mean (SD)]	1.379	0.539	0.648	1.422	2.095	2.883
Pain phenotype [PDQ7; NRS-35; mean (SD)]	0.599	0.216	0.182	0.57	0.961	1.151
Harmonized standard deviation scores (hSDS): mean	1.798	0.109	1.241	2.412	3.398	3.514

Within-group effect sizes quantifying treatment-associated differences between baseline and week-3 768 harmonized standard deviation scores (hSDS) across biopsychosocial severity groups. Cohen's d values are presented to characterize the magnitude of within-group treatment-associated changes, with larger values indicating greater standardized improvement.

In the highest severity groups, improvements exceeding one standard deviation occurred considerably more frequently than in lower severity categories.

Baseline-adjusted analyses

To further evaluate treatment-associated changes while accounting for baseline differences, analyses of covariance were performed.

The first model examined Δ-hSDS as dependent variable and included severity group, BL-hSDS, and the interaction between BL-hSDS and severity group as predictors (Table 5a).

Table 5: Baseline-adjusted analysis of covariance model for treatment-associated changes (Δ-hSDS).

A: Model effects

Predictor	df	F	p-value	Partial η^2
Severity group	4	15.64	<0.001	0.058
Baseline z-score (BL-ZS)	1	570.32	<0.001	0.008
Severity group × BL-ZS	4	18.25	<0.001	0.041
Overall model	9	1350.28	<0.001	R^2=0.569

B: Estimated marginal means (adjusted for BL-ZS)

Severity group	Adjusted mean	95% CI
1	0.699	0.561–0.837
2	0.823	0.795–0.851
3	0.943	0.928–0.957

4	0.991	0.934–1.048
5	0.727	0.343–1.111

(Annotation: Estimated marginal means were calculated at BL-ZS = 0.)

A: Results of the analysis of covariance model with treatment-associated changes in harmonized standard deviation scores (Δ-hSDS) as dependent variable and biopsychosocial severity group (BPSG), baseline harmonized standard deviation scores (BL-hSDS), and their interaction included as predictors. Model effects are presented as F-values, p-values, and partial eta-squared (η^2).

B: Estimated marginal means (EMMs) for Δ-hSDS adjusted for baseline values and calculated at BL-779 hSDS = 0. Corresponding 95% confidence intervals are shown.

The overall model explained a substantial proportion of variance in Δ-hSDS (R^2 = 0.569). Severity group showed a statistically significant association with Δ-hSDS ($p < 0.001$), with a moderate effect size (partial η^2 = 0.058). BL-hSDS was also significantly associated with treatment-associated change, although with a comparatively smaller effect size (partial η^2 = 0.008).

Furthermore, a statistically significant interaction between BL-hSDS and severity group was observed (partial η^2 = 0.041), indicating that the relationship between baseline status and standardized treatment-associated change differed across severity categories.

Estimated marginal means adjusted for baseline values are presented in Table 5b. Adjusted Δ-hSDS values ranged from 0.699 to 0.991 across severity groups. Compared with unadjusted analyses, differences between groups appeared less pronounced following baseline adjustment.

Week-3 outcome analyses

A second analysis of covariance examined W3-hSDS as dependent variable and included severity group and BL-hSDS as predictors (Table 6a).

The overall explained variance of this model was lower (R^2 = 0.097). BL-hSDS represented the dominant predictor of week-3 outcome values (partial η^2 = 0.517), whereas severity group showed a comparatively small but statistically significant effect (partial η^2 = 0.008).

Estimated marginal means adjusted for baseline values increased progressively across severity groups and ranged from 0.743 to 1.108 (Table 6b).

Table 6: Baseline-adjusted analysis of covariance model for standardized week-3 outcomes (W3-hSDS).

A: Model effects

Predictor	df	F	p-value	Partial η^2
Severity group	4	19.67	<0.001	0.008
Baseline z-score	1	9854.73	<0.001	0.517
Overall model	5	197.77	<0.001	R^2=0.097

B: Estimated marginal means

Severity group	Adjusted mean	95% CI
1	0.743	0.693–0.792
2	0.888	0.866–0.910
3	0.959	0.946–0.973
4	1.038	1.010–1.066
5	1.108	1.023–1.192

A: Results of the analysis of covariance model with standardized week-3 outcome values (W3-hSDS) as dependent variable and biopsychosocial severity group (BPSG) together with baseline harmonized standard deviation scores (BL-hSDS) included as predictors. Model effects are presented as F-values, p-values, and partial eta-squared (η^2).

B: Estimated marginal means (EMMs) for W3-hSDS adjusted for baseline values and calculated at BL-hSDS = 0. Corresponding 95% confidence intervals are shown.

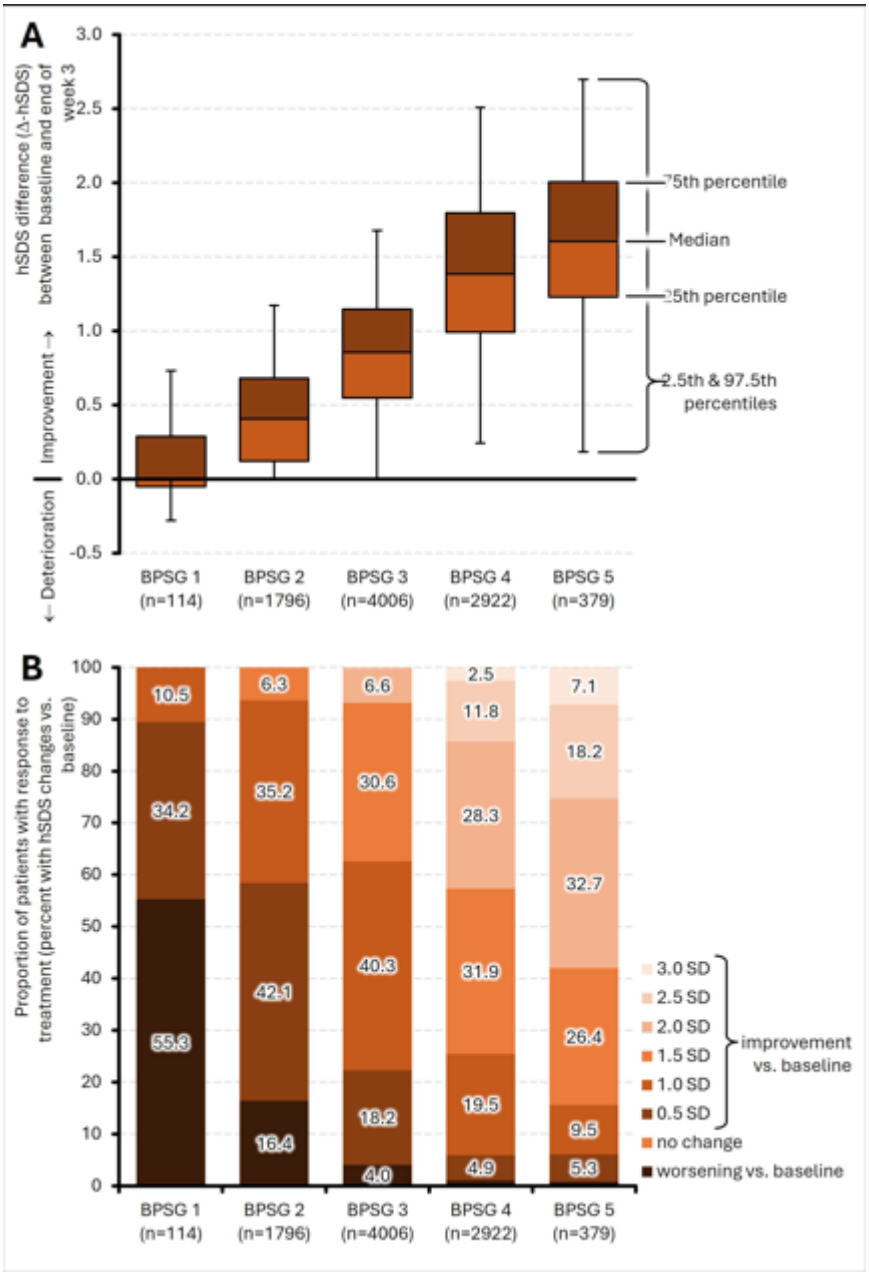

Figure 4: Standardized treatment response according to biopsychosocial severity group.
(A) Boxplots of treatment-associated changes in harmonized standard deviation scores (Δ-hSDS) across biopsychosocial severity groups (BPSG 1–5). Boxes represent interquartile ranges, horizontal lines indicate median values, whiskers represent the 95th percentile range, and circles indicate outliers. (B) Distribution of responder categories according to magnitude of standardized improvement. Response categories are expressed as proportions of patients achieving predefined thresholds ranging from deterioration/no change to ≥3 standard deviations of improvement.

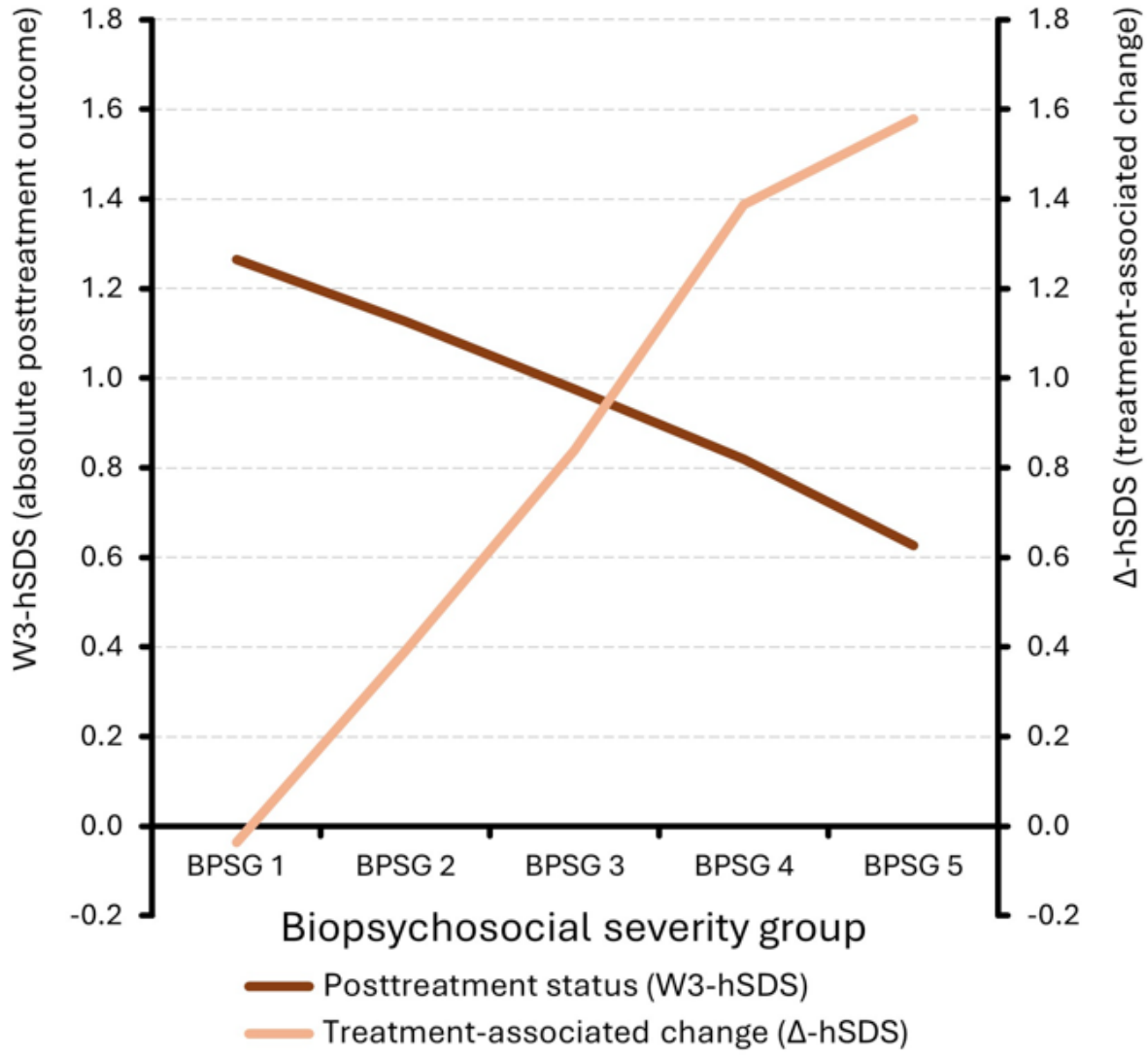

Figure 5: Conceptual summary of the relationship between baseline severity, treatment-associated change, and postreatment outcome.
Illustrative summary based on observed harmonized standardized outcomes across biopsychosocial severity groups (BPSG 1–5). Increasing baseline severity was associated with progressively larger treatment-associated improvements (Δ-hSDS), whereas absolute postreatment outcome levels (W3- hSDS) remained lower in more severely affected patients. The figure highlights the distinction between treatment response and postreatment status and summarizes implications for severity- based treatment allocation.

Discussion

Principal findings

The present study examined the relationship between multidimensional biopsychosocial baseline severity and short-term treatment-associated changes following structured interdisciplinary outpatient multimodal pain therapy in a large cohort of patients with persistent spinal pain. Across all analyses, a consistent relationship between baseline severity and treatment-associated change was observed. Standardized improvements increased progressively across severity categories, with the largest treatment-associated changes and responder proportions occurring in patients with severe and very severe baseline impairment.

Contrary to common assumptions that high biopsychosocial burden may limit suitability for outpatient treatment pathways, substantial treatment-associated improvements were not restricted to mildly affected patients. This pattern was observed consistently across unadjusted analyses, responder distributions, and baseline-adjusted models. In contrast, absolute posttreatment outcome levels remained associated with baseline status, indicating that treatment-associated change and posttreatment status represent distinct outcome dimensions.

One potential explanation for this finding may relate to the individualized and adaptive structure of the integrated care model. Within the BV-S framework, treatment allocation was based on

interdisciplinary bio-psycho-social assessment and shared decision-making procedures [12].

Treatment intensity and therapeutic components were aligned with clinical complexity and psychosocial burden. Patients with greater multidimensional impairment may therefore have received interventions more closely matched to their individual treatment needs, potentially contributing to greater treatment responsiveness despite more severe baseline presentations.

Interpretation of the severity-response relationship

The observed association between higher baseline burden and larger standardized treatment-associated improvements is clinically relevant and requires careful interpretation. Patients with severe multidimensional impairment generally exhibit greater limitations across several domains, including pain-related disability, psychological distress, maladaptive beliefs, and social participation [10,16,17]. Consequently, these patients may possess a larger range of measurable changes across multidimensional outcome constructs.

Interdisciplinary multimodal treatment specifically targets these interacting dimensions simultaneously rather than focusing on isolated symptom reduction. Therefore, patients with broader biopsychosocial impairment may be particularly responsive to interventions addressing multiple determinants of pain chronification in parallel [5-7].

At the same time, larger changes in patients with unfavorable baseline status cannot automatically be interpreted as direct evidence of stronger treatment effects. Statistical mechanisms including regression to the mean and scale-related constraints likely contribute to the observed pattern [19,20]. Patients with extreme baseline values are more likely to demonstrate movement toward average values on repeated measurements independent of treatment exposure. Likewise, ceiling effects may reduce measurable improvement in patients with comparatively favorable baseline status [24].

The pronounced inverse association between BL-hSDS and Δ-hSDS observed in the present analyses illustrates these phenomena. However, several observations argue against interpreting the findings solely as statistical artifacts. The observed pattern was consistent across multiple analytical approaches including standardized change scores, responder analyses, and baseline-adjusted models.

Furthermore, clinically meaningful improvements remained observable even after accounting for baseline differences. In addition, the magnitude of observed standardized improvements suggests that these changes are unlikely to represent trivial statistical fluctuations alone. Improvements approaching or exceeding one standard deviation generally indicate substantial shifts across multiple domains and therefore likely reflect clinically meaningful treatment-associated changes rather than isolated measurement variability [21,22].

7Change and outcome represent different clinical constructs

An important finding of the present analysis is the distinction between treatment-associated change and absolute posttreatment outcome.

Patients with severe baseline burden demonstrated the largest improvements yet remained at lower absolute week-3 levels compared with less impaired patients. This observation indicates that large treatment-associated changes do not necessarily result in complete normalization of clinical status. This distinction is highly relevant in clinical practice. Treatment success may be evaluated either according to the magnitude of improvement or according to the achievement of a predefined health state. These perspectives are not interchangeable and may lead to different conclusions regarding effectiveness [25-28].

A severely affected patient may experience clinically meaningful functional gains and substantial reductions in multidimensional burden while remaining symptomatic after treatment. Conversely, patients entering treatment with relatively favorable baseline status may show only limited measurable improvement despite comparatively good outcome levels.

The present findings therefore support simultaneous consideration of both change-based and status- based outcome measures when evaluating interdisciplinary pain programs. The overall relationship between baseline severity, treatment-associated change, and posttreatment status is summarized in Figure 5. While patients with greater biopsychosocial burden demonstrated larger treatment- associated improvements, absolute posttreatment outcomes remained associated with baseline status. This distinction may have important implications for interpretation of treatment effectiveness and severity-based treatment allocation.

Implications for outpatient versus inpatient treatment allocation

The present findings have direct implications for stepped-care concepts in persistent spinal pain. In routine care, patients with high levels of biopsychosocial burden are frequently considered candidates for escalation of care and more intensive treatment settings, including inpatient multimodal therapy. High levels of disability, psychological distress, chronicity, and social impairment are often interpreted as indicators of reduced suitability for outpatient treatment approaches [10,11].

In addition, treatment delivered within patients' real-world social and occupational environments may facilitate transfer of therapeutic strategies into everyday life [12,29]. Ambulatory care structures may preserve social participation, reinforce self-management capacities, and support implementation of behavioral changes under everyday conditions. These characteristics may be particularly relevant for patients with pronounced functional and psychosocial burden [29].

The present data do not support such a simplified interpretation. The highest severity categories demonstrated the largest standardized improvements and contained substantial proportions of patients achieving clinically meaningful response levels.

Importantly, this finding should not be interpreted as evidence that inpatient treatment is unnecessary or inferior. The present study did not include direct comparison between outpatient and

inpatient treatment settings and therefore cannot determine comparative effectiveness. Likewise, it remains possible that certain patient subgroups may derive additional benefit from more intensive treatment environments.

However, the findings suggest that baseline severity alone should not automatically be interpreted as evidence against outpatient multimodal treatment or as a sole indication for inpatient care. Structured interdisciplinary outpatient programs appear capable of producing substantial short-term benefit even among patients with pronounced biopsychosocial burden. Taken together, our findings challenge simplified assumptions that greater baseline burden necessarily predicts limited outpatient treatment responsiveness.

Methodological implications of harmonized standardized outcomes

The use of harmonized standard deviation scores represented a central methodological component

of the present analysis. The included instruments differ considerably regarding scale range, orientation, and psychometric properties. Direct aggregation of raw score changes or percentage differences across these measures would therefore have introduced substantial interpretative limitations.

Harmonized standardization enabled expression of multidimensional outcomes on a common metric and facilitated integration across domains [21-23]. This approach also permitted direct comparison of treatment-associated changes between severity groups.

At the same time, composite standardized outcomes have limitations. Although they improve comparability across instruments, they may obscure domain-specific response patterns. Patients with similar global scores may improve in different dimensions such as disability, psychological distress, or self-efficacy [17,18,28].

Future studies should therefore complement multidimensional global outcomes with analyses examining specific response domains.

Strengths and Limitations

The present study has several important strengths. First, the analysis is based on a large multicenter cohort comprising 9,217 patients treated within a structured nationwide interdisciplinary integrated care network. The large sample size enabled stable estimation across five predefined severity groups and allowed detailed analyses of treatment response patterns across the full spectrum of biopsychosocial burden. Another important strength lies in the treatment setting itself. Patients were managed within a highly structured interdisciplinary integrated care architecture including standardized interdisciplinary assessment procedures, adaptive treatment allocation, and longitudinal quality assurance structures. This framework increases ecological validity and reflects routine clinical decision-making under real-world conditions.

Second, the study population represents a clinically highly relevant and challenging patient group. All patients underwent interdisciplinary second-opinion assessment prior to planned spinal surgery and subsequently participated in conservative treatment following non-confirmation of surgical indication. This population frequently presents with complex clinical characteristics including long- standing pain histories, substantial disability, psychological distress, and pronounced biopsychosocial impairment. Findings derived from such populations are of direct relevance for treatment allocation and healthcare planning.

Third, baseline severity was not operationalized using a single dimension such as pain intensity or disability alone but through a multidimensional biopsychosocial severity model integrating functional, psychological, social, and chronicity-related domains. This approach more closely reflects contemporary conceptualizations of persistent pain and allowed examination of treatment response across clinically meaningful severity profiles.

Fourth, harmonized standard deviation scores (hSDS) were used to integrate outcomes across instruments with substantially different scaling properties. By harmonizing directionality and standardizing values relative to baseline distributions, multidimensional treatment-associated changes could be expressed on a common metric. The combination of standardized outcomes, responder analyses, graphical approaches, and baseline-adjusted models provided a comprehensive framework for evaluating treatment-associated change.

However, several limitations should also be acknowledged.

The most important limitation is the observational design without a randomized comparison group. Consequently, treatment-associated improvements cannot be attributed exclusively to the intervention itself. Natural symptom variation, contextual treatment effects, expectation effects, spontaneous improvement, and additional non-specific factors may have contributed to observed changes.

Second, the analyses demonstrated a strong association between baseline status and treatment-associated change. Therefore, statistical mechanisms including regression to the mean and scale-related effects must be considered. Patients with extreme baseline values generally have greater opportunity for measurable movement toward average values at repeated assessment. Although the present analyses explicitly modeled and visualized these relationships through baseline-adjusted approaches and interaction analyses, complete separation of statistical and treatment-related effects is not possible in an observational setting.

Third, although harmonized standardized outcomes improve comparability across domains, they necessarily represent a composite construct. Patients with similar global hSDS changes may have improved in different dimensions of functioning and distress. Consequently, the present analyses provide a multidimensional summary of response but do not replace detailed domain-specific evaluation.

Fourth, the multidimensional severity model itself requires further validation. Although score construction was rule-based, clinically plausible, and intentionally designed to represent biopsychosocial burden, all domains contributed equally to the final score. It remains possible that individual domains differ in prognostic rel-

evance and alternative weighting procedures may further improve predictive performance.

Fifth, outcome assessment was limited to the end of the three-week intervention period. Therefore, the present analysis addresses short-term treatment-associated change only. Long-term follow-up data are required to determine whether the observed improvements remain stable over time and whether outpatient treatment influences later healthcare utilization, surgical rates, or need for inpatient treatment.

Sixths, the biopsychosocial severity model was derived and evaluated within the same observational cohort. External validation in independent patient populations will therefore be required before broader implementation of the proposed severity classification can be recommended.

Finally, the study was performed within a highly structured integrated care pathway and included patients following specialized interdisciplinary second-opinion assessment. Therefore, generalizability to less structured outpatient environments or healthcare systems with different organizational characteristics may be limited. Furthermore, although treatment was embedded within a broader integrated care architecture including structured refresher interventions and longitudinal follow-up components, the present analyses focused exclusively on short-term treatment-associated changes at completion of the initial three-week intervention. Potential contributions of subsequent care components were therefore not evaluated.

Conclusions

In this large multicenter cohort of patients with persistent spinal pain, structured interdisciplinary outpatient multimodal pain therapy was associated with substantial short-term improvements across the full spectrum of multidimensional biopsychosocial baseline severity. Patients with severe and very severe baseline impairment demonstrated pronounced treatment-associated changes and substantial responder rates despite high levels of initial burden.

Although treatment-associated changes remained strongly related to baseline status and require cautious interpretation in light of regression-to-the-mean and scale-related effects, the findings consistently indicate that clinically meaningful improvement is not restricted to mildly affected patients.

These results suggest that high biopsychosocial burden alone should not be interpreted as evidence against outpatient interdisciplinary multimodal treatment or as an automatic indication for inpatient care. Rather, structured outpatient programs may represent an effective component of stepped-care approaches even for patients with substantial baseline impairment.

Future controlled studies with long-term follow-up are needed to determine which patient subgroups benefit most from outpatient versus inpatient treatment pathways and to further refine severity- based treatment allocation strategies.

Conflict of Interest

The authors are founders and owners of Integrated Managed Care GmbH (IMC), which developed and coordinates the BV-S integrated care framework described in this article.

Acknowledgments

The authors would like to thank all participating physicians, therapists, and clinical staff within the interdisciplinary outpatient pain treatment network for their contribution to patient care and data documentation within the German Pain e-Registry (GPeR). The authors also acknowledge the patients who consented to the use of their anonymized data for health services research.

References

1. Hartvigsen J, Hancock MJ, Kongsted A, Louw Q, Ferreira ML, et al. (2018) What low back pain is and why we need to pay attention. Lancet 391(10137): 2356-2367.
2. GBD 2019 Diseases and Injuries Collaborators (2020) Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet. Oct 17;396(10258):1204-1222. Erratum in: Lancet 396(10262): 1562.
3. Jensen MC, Brant-Zawadzki MN, Obuchowski N, Modic MT, Malkasian D, et al. (1994) Magnetic resonance imaging of the lumbar spine in people without back pain. N Engl J Med 331(2): 69-73.
4. Brinjikji W, Luetmer PH, Comstock B, Bresnahan BW, Chen LE, et al. (2015) Systematic literature review of imaging features of spinal degeneration in asymptomatic populations. AJNR Am J Neuroradiol 36(4): 811-6.
5. Pfingsten M, Arnold B, Böger A, Brinkschmidt T, Casser HR, et al. (2019) Sektorenübergreifende interdisziplinäre multimodale Schmerztherapie : Empfehlungen zu Struktur- und Prozessparametern der Ad-hoc-Kommission „Interdisziplinäre Multimodale Schmerztherapie" der Deutschen Schmerzgesellschaft e. V [Cross-sectoral interdisciplinary multimodal pain therapy: Recommendations on structural and process parameters of the ad hoc commission "Interdisciplinary Multimodal Pain Therapy" of the German Pain Society (Deutsche 608 Schmerzgesellschaft e. V.)]. Schmerz 33(3): 191-203.
6. Nicholas MK, George SZ (2011) Psychologically informed interventions for low back pain: an update for physical therapists. Phys Ther 91(5): 765-76.
7. Kamper SJ, Apeldoorn AT, Chiarotto A, Smeets RJ, Ostelo RW, et al. (2014) Multidisciplinary biopsychosocial rehabilitation for chronic low back pain. Cochrane Database Syst Rev 2014(9): CD000963.
8. Von Korff M, Ormel J, Keefe FJ, Dworkin SF (1992) Grading the severity of chronic pain. Pain 50(2):133-149.
9. Main CJ, Sowden G, Hill JC, Watson PJ, Hay EM (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(2): 110-6.
10. Gatchel RJ, McGeary DD, McGeary CA, Lippe B (2014) Interdisciplinary chronic pain management:past, present, and future. Am Psychol 69(2):119-30.
11. Vlaeyen JWS, Linton SJ (2000) Fear-avoidance and its consequences in chronic musculoskeletal pain: a state of the art. Pain 85(3): 317-332.
12. Überall MA, Nolte T, Müller-Schwefe GHH, Kletzko H. BV-S
13. Lurie JD, Weinstein JN (2001) Shared decision-making and the orthopaedic workforce. Clin Orthop Relat Res (385): 68-75.
14. Mannion AF, Mariaux F, Reitmeir R, Fekete TF, Haschtmann D, et al. (2020) Development of the "Core Yellow Flags Index" (CYFI) as a brief instrument for the assessment of key psychological factors in patients undergoing spine surgery. Eur Spine J 29(8): 1935-1952.

15. Mansell G, Hill JC, Main CJ, Von Korff M, van der Windt D (2017) Mediators of Treatment Effect in the Back In Action Trial: Using Latent Growth Modeling to Take Change Over Time Into Account. Clin J Pain 33(9): 811-819.

16. Turk DC, Rudy TE (1987) IASP taxonomy of chronic pain syndromes: preliminary assessment of reliability. Pain 30(2):177-189.

17. Dworkin RH, Turk DC, Wyrwich KW, Beaton D, Cleeland CS, et al. (2008) Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. J Pain 9(2): 105-21.

18. Barnett AG, van der Pols JC, Dobson AJ (2005) Regression to the mean: what it is and how to deal with it. Int J Epidemiol 34(1): 215-20. Erratum in: Int J Epidemiol. 2015 Oct;44(5):1748.

19. Twisk JW, de Vente W (2008) The analysis of randomised controlled trial data with more than one follow-up measurement. A comparison between different approaches. Eur J Epidemiol 23(10): 655-60.

20. Norman G (2003) The effectiveness and effects of effect sizes. Adv Health Sci Educ Theory Pract 8(3): 183-7.

21. Sullivan GM, Feinn R (2012) Using Effect Size-or Why the P Value Is Not Enough. J Grad Med Educ Ep 4(3): 279-82.

22. Ware J Jr, Kosinski M, Keller SD (1996) A 12-Item Short-Form Health Survey: construction of scales and reliminary tests of reliability and validity. Med Care 34(3):220-33.

23. Wang L, Zhang Z, McArdle JJ, Salthouse TA (2009) Investigating Ceiling Effects in Longitudinal Data Analysis. Multivariate Behav Res 43(3): 476-496.

24. Deyo RA, Mirza SK, Martin BI (2002) Back pain prevalence and visit rates: estimates from U.S. national surveys Spine (Phila Pa 1976) 31(23): 2724-7.

25. Farrar JT, Young JP Jr, LaMoreaux L, Werth JL, Poole MR (2001) Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. Pain 94(2):149-158.

26. Tubach F, Ravaud P, Martin-Mola E, Awada H, Bellamy N, et al. (2012) Minimum clinically important improvement and patient acceptable symptom state in pain and function in rheumatoid arthritis, ankylosing spondylitis, chronic back pain, hand osteoarthritis, and hip and knee osteoarthritis: Results from a prospective multinational study. Arthritis Care Res (Hoboken) 64(11): 1699-707.

27. Weldring T, Smith SM (2013) Patient-Reported Outcomes (PROs) and Patient-Reported Outcome Measures(PROMs). Health Serv Insights 6: 61-8.

28. Überall MA, Nolte T, Müller-Schwefe GHH, Kletzko H. CPGS.