

Cost-Utility Analysis of a Chronic Back Pain Multidisciplinary Biopsychosocial Rehabilitation (MBR) compared to Standard Care for Privately Insured in Germany

Martin Hochheim (✉ martinhochheim@gmail.com)

Medizinische Hochschule Hannover <https://orcid.org/0000-0001-9572-3022>

Philipp Ramm

Generali Health Solutions GmbH

Max Wunderlich

Generali Health Solutions GmbH

Volker Amelung

Medizinische Hochschule Hannover

Research article

Keywords: Chronic Back Pain, Cost-effectiveness, Difference-in-Difference, Korff's Chronic Pain Grade Questionnaire, Multidisciplinary biopsychosocial rehabilitation (MBR), Programme Evaluation

Posted Date: June 29th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-659197/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Introduction: Multidisciplinary biopsychosocial rehabilitation (MBR) is highly recommended to treat chronic lower back pain (CLBP). However, its economic benefit remains to be clearly demonstrated.

Objective: To analyse the effect of a 12-month MBR with behaviour-change coaching and device-supported exercise on direct medical costs, sick leave and health-related quality of life (HRQOL) at 24 months.

Methods: A cohort of privately insured in Germany was evaluated using administrative and trial data. After removing dissimilarities in characteristics between MBR and control via propensity score matching, treatment effects were calculated with a Difference-in-Difference approach.

Results: The base-case analysis comprised 112 MBR participants and 111 members in the standard care group. With an incremental cost-effectiveness ratio (ICER) of €8,428 per quality-adjusted life year (QALY) gained, the intervention was classified as cost-effective. Economically unaccounted sick leave due to back pain (BP) in the last six months was reduced by 17.5 days ($p = 0.001$) in the MBR group. Development of HRQOL was positive for the MBR (0.046, $p = 0.026$). Subgroup analysis with major impaired participants demonstrated the possibility of a dominant intervention with an ICER of - €6,861 per QALY. Savings were driven by a reduction in BP specific costs by - €1,733 ($p = 0.035$). Difference in sick leave was 27 days ($p = 0.006$) in favour of the MBR group.

Conclusion: This is the first cost-utility study with combined data from a private health insurer and a controlled trial to demonstrate that long term MBR is cost-effective in the treatment of CLBP. Subgroups with major impairment benefit more from the intervention than participants with minor impairment. MBR significantly reduces sick leave in all participants making it a profitable intervention from society's point of view.

Introduction

Lower back pain (BP) is the leading cause of years lived with disability (YLD) worldwide [1]. For those affected, frequent absences from work and severe pain-related disability are common. BP causes high direct and very high indirect costs [2]. Estimates of the economic burden of BP range from annual costs of AU\$9.17 billion in Australia to \$91 billion in the USA [3, 4]. In Germany, costs are estimated to be up to €49 billion per year [5].

Chronic lower back pain (CLBP) is an urgent global public health concern [6]. It is often conceptualised as a biopsychosocial problem, i.e. a complex and dynamic interaction between physical, psychological and social elements [7]. To treat CLBP, multidisciplinary biopsychosocial rehabilitation (MBR) – a combination of physical and behavioural or social components - is recommended in clinical guidelines [8].

However, the actual care often does not correspond to clinical recommendations. A frequent oversupply of low-value care and undersupply of high-value care exists. Passive therapies such as massage or MRI are frequently provided, although they are associated with high costs and little benefit. [8].

Even if guideline-compliant treatment is carried out, it is not clear whether it is cost-effective [9]. Evidence on the cost-effectiveness of structured long-term treatments is still scarce, which is why it has been rated as a research priority [10]. From a health care provider's perspective, it is important to take the health economic efficiency into account when deciding on which treatment to offer.

Several papers have shown cost-effectiveness of short-term interdisciplinary multimodal pain therapy (IMPT) [11–13]. However, most studies focus either on cost development [11] or medical effectiveness [12] but fail to retrieve both from the same study population. Due to their high treatment dosage, duration and intensity, IMPT treatments are costly and difficult to access [14].

Studies focusing on the cost-benefit of medium to long-term standardised treatment programmes for non-specific BP are very diverse. Therapies (*physical, information & education, manual therapy, combined physical & psychological interventions*), comparators (*outpatient physiotherapy, usual care, usual care & component, self-care advice*), periods considered (*3 - 120 months*), perspective adopted (*individual, society, health care provider*), sources of costs (*administrative vs. calculated from participants reporting*) and target group (*acute vs. chronic*) differ vastly.

Herman et al. attempted to sort out this diversity and calculated a Markov model comparing each of the 27 investigated forms of therapy [15, 16]. Respective treatment costs, differences in BP specific healthcare costs, productivity costs, total costs for society and provider, as well as the quality-adjusted life years (QALY) between usual care and the form of intervention were presented. The average cost of treatment was \$420. The cheapest intervention was a group-based cognitive behavioural therapy [17] at \$71 per participant, the most expensive intervention was an active trunk exercise [18] at \$1,077.

The QALY gain compared to standard care across all programmes was 0.017. Active training therapies had the highest QALY gain (flexion-distraction [18] = 0.033, active trunk exercise [18] = 0.033, yoga [19] = 0.048). Traditional Chinese medicine acupuncture [20] and spinal manipulation [21] had the lowest QALY gain (both 0.004). The mean incremental cost-effectiveness ratio (ICER) was \$3591 per QALY. Sixteen therapies were dominant (i.e. less back-specific costs with simultaneously higher benefits) compared to standard care. According to the model, yoga has the best cost-benefit ratio with a reduction in back-specific costs of \$1,136 and a simultaneous QALY gain of 0.048 [19]. Spinal manipulation [18] has the worst cost-benefit ratio with a QALY gain of 0.004 at an additional cost of \$457 and a resulting ICER of \$114,250. The only MBR intervention considered [20] has a QALY gain of 0.015 and a cost reduction of \$172 resulting in an ICER of - \$11,466.

However, as only BP specific and not all direct costs were taken into account for the ICER calculation, the results must be interpreted cautiously. CLBP as a biopsychosocial problem has various causes. They are often associated with psychological disorders, multiple medical issues, including obesity, smoking and

lack of exercise [22, 23]. It therefore seems to be too narrow to consider only BP specific costs in the economic evaluation.

The working groups around Groessl [24], Lambeek [25] and Williams [26] also analysed structured interventions for CLBP and compared them with usual care. They are not included in the model of Herman et al. and actually used all direct health care costs for the QALY calculation.

Groessl et al. [24] investigated a 12-week long progressive yoga intervention on a total of 150 participants. They reported a QALY gain of 0.043, additional healthcare cost of \$193 with a resulting ICER of \$4,488.

Lambeek et al. [25] reported on a workplace intervention consisting of a graded activity programme with cognitive behavioural principles for sick-listed patients with CLBP. Sixty-six participants received the intervention, 68 were treated according to usual care. After 12 months, a QALY gain of 0.09 and additional direct costs of £217 were described, leading to an ICER of £2,411.

Williams et al. [26] analysed the effectiveness of a healthy lifestyle intervention consisting of a brief advice, a clinical consultation and referral to a 6-month telephone-based coaching service. The intervention participants (n=80) had a QALY gain of 0.02 and increased total direct costs from a provider perspective of AU\$386 compared to usual care (n=79), resulting in an ICER of AU\$19,036.

Müller et al. [27] presented a cost-effective study for a long-term multimodal intervention with a device-oriented exercise approach and broad access options for a German population. The intervention group consisted of 1,829 participants and the usual care control group of 495 members. After a follow-up time of two years, they saw a reduction of direct medical costs of €763 and BP specific cost savings of €239. The biggest effect could be seen in the most affected group according to chronicity grades with savings of €4,535. The authors did not provide information about the development of the QALY. As neither a psychological nor a social dimension was included in their intervention analysed, it cannot be classified as an MBR. To the author's knowledge, no published cost-effective analysis for a long term MBR in Germany exists so far.

Four systematic reviews researching cost-benefit studies in the treatment of CLBP [9, 28–30] exist. All of them called for more high-quality site-specific studies to reduce the uncertainty of cost effects in BP therapies, asked for observation periods longer than 12 months and recommended using high reporting standards.

This paper addresses the gap and fulfils the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) criteria in analysing whether a 12-month outpatient MBR intervention consisting of behaviour-change coaching and device-supported exercise with low entrance barriers can prove cost-effective compared to standard of care in patients with CLBP in a private health insurance (PHI) setting in Germany.

Routine data from the PHI provider, Generali Deutschland Krankenversicherung AG (Generali Germany Health Insurance, formerly known as "Central Krankenversicherung") is evaluated. Since 2014, Generali has been running a one-year MBR for people with CLBP. The intervention was accompanied by a controlled trial in Zelen's -Design, registered at the German Clinical Trials Register under DRKS00015463. Results showed medical effectiveness in the 12-month follow-up [31, 32]. To find out whether the intervention is also cost-effective, inpatient as well as outpatient costs, the development of sick days due to BP and the overall health status of the participants enrolled in the aforementioned study were collected and analysed in detail.

By isolating BP specific costs, judgements about long-term effectiveness on BP outcomes can be deducted. Incremental costs in addition to the effectiveness of the MBR are calculated, complemented by insights about the effect on sick leave due to BP. This is a first-time cost-utility evaluation - in accordance with the CHEERS statement - of a long term MBR for patients with CLBP in Germany.

Methods

Target population and subgroups

The present economic evaluation reuses data from a two-arm randomised controlled evaluation study in parallel-group Zelen's design for which recruitment took place between April and October 2015 [31, 32] and amends it with administrative direct health costs data. Searches in the database of the PHI identified adults with administrative indications of CLBP. The full inclusion criteria can be found elsewhere [32].

One of the main results of the previously run effectiveness analysis stated the effect of the intervention being highly dependent on the level of impairment due to BP [32]. Therefore we divided participants into subgroups based on their overall result in the *Chronic Pain Grade Questionnaire* [33, 34]. Two parameters were used to classify BP severity levels: the characteristic pain intensity (score 0-100) as an average of the current, average and maximum pain intensity, and the pain-related impairment (0-6 points), calculated from the number of impairment days and the extent of the impairment experienced in everyday life, leisure and work. These led to four hierarchical Graded Chronic Pain Grades (GCPS): Grade I, low disability-low intensity; Grade II, low disability-high intensity; Grade III, high disability-moderately limiting; and Grade IV, high disability-severely limiting [33].

In the following analysis, the grades were combined: Grade I and Grade II as *minor impaired* (functional chronic pain) and Grades III and IV as *major impaired* (dysfunctional chronic pain) due to BP.

Table 1 shows the characteristics of the population before and after data processing and matching (see analytic methods). It can be observed that after the processing, participants in the intervention group (IG) and in the control group (CG) are almost equally distributed. Group size (n: IG = 112, CG = 111), as well as characteristics, are very similar in both groups at the beginning of the study period so that the costs incurred can be compared well. The mean age is 55.19 years. A sex distribution of 35 % female and 65 % male participants is present. Of the study population, 53 % are categorised as *minor* and 47 % as *major*

impaired by BP. According to Keele STarT Back Screening Tool [35] 55 % are of low, 33 % of medium and 12 % of high risk of persisting disabling symptoms.

Based on the weighted Charlson Comorbidity Index Score (CCI) [36], used to classify comorbidities in the likelihood of mortality/high resource use, it can be stated that the groups are comparable at baseline.

Table 1 - Baseline Characteristics of IG and CG before and after data processing						
	Before			After		
	IG	CG	p	IG	CG	P
<i>n</i>	189	254		112	111	
<i>Age (mean (SD))</i>	53.86 (8.13)	54.14 (8.65)	0.725	55.68 (7.34)	54.69 (8.45)	0.354
<i>Gender = Female (%)</i>	70 (31.7)	103 (40.6)	0.072	39 (34.8)	40 (36.0)	0.960
<i>GCPS (%)</i>			0.062			0.948
<i>I</i>	64 (33.9)	115 (45.3)		41 (36.6)	42 (37.8)	
<i>II</i>	26 (13.8)	36 (14.2)		18 (16.1)	19 (17.1)	
<i>III</i>	52 (27.5)	49 (19.3)		28 (25.0)	24 (21.6)	
<i>IV</i>	47 (24.9)	54 (21.3)		25 (22.3)	26 (23.4)	
<i>STarT-Back (%)</i>			0.068			0.942
<i>1</i>	100 (52.9)	161 (63.4)		62 (55.4)	60 (54.1)	
<i>2</i>	64 (33.9)	71 (28.0)		36 (32.1)	28 (34.2)	
<i>3</i>	25 (13.2)	22 (8.7)		14 (12.5)	13 (11.7)	
<i>EQ-5D (mean (SD))</i>	0.61 (0.18)	0.63 (0.20)	0.253	0.61 (0.18)	0.60 (0.19)	0.78
<i>CCI</i>			0.661			0.6
<i>0</i>	76 (49.4)	127 (50)		56 (50)	47 (42.3)	
<i>1-2</i>	54 (35.1)	97 (38.2)		40 (35.7)	46 (41.4)	
<i>3-4</i>	20 (13.0)	23 (9.1)		14 (12.5)	14 (12.6)	
<i>>=5</i>	4 (2.6)	7 (2.8)		2 (1.8)	4 (3.6)	

Table 1 - Baseline Characteristics of IG and CG before and after data processing

Setting and location

There is a division of health care in Germany. About 90 % of the population is a member of the statutory health insurance (SHI), and 10 % is privately health insured. The latter implicates a comprehensive health insurance plan dealing as a substitute for the mandatory SHI. Usually, members of the PHI are self-employed, civil servants or employees with a salary above the compulsory insurance threshold (currently €64,350). In general, members of the PHI belong to a higher socio-economic class.

Due to the nature of the system ('lock-in effect'), there is a low fluctuation within the membership [37]. The insurer has an interest in healthy members and therefore considers investments over a longer time period. The payer is therefore interested in optimal, evidence-based care for its insurees. Unlike other chronic diseases, still no uniform, guideline-based disease management programme (DMP) for BP exists in Germany [38]. Individual insurance companies have developed and implemented different approaches, which are however rarely evaluated scientifically [32, 39]. As an innovation driver from payer to health care partner, a PHI company in Germany can pave the way and test the effectiveness and efficiency of a structured treatment programme [40]. This could be a good blueprint for the introduction of a DMP for BP in the SHI.

This study was set within the Generali Germany Health Insurance. In 2019, Generali listed 308,088 fully insured members and 1,431,522 with supplementary insurance [41] and was one of the largest PHI in Germany. Only fully insured members were eligible to participate in the offered health programme.

Study perspective

The analysis was conducted from the payer's perspective. Total health costs were evaluated, as well as costs that could be attributed to chronic back pain (ICD M40-M54 diagnosis).

The costs were divided into outpatient and inpatient costs. In order to achieve better comparability with the statutory system, additional elective services (e.g. one or two-bedroom supplement) and the entire costs of dental care treatment were excluded. Included were costs from the following areas: General hospital services, GP and specialist care, medicines, remedies, alternative practitioners (e.g. chiropractor), aids and private medical treatment.

In a PHI setting, reimbursement of the health care bills depends on the respective tariff. The study population consisted of fully insured participants as well as policyholders eligible for governmental aid. Therefore, the bill amount was used as the costs and not the amount of refund paid by the insurance. Thus, the actual costs were compared with each other without taking into account which payer (health insurance, subsidy or individual supplementary) reimbursed the costs.

In order to allow further analysis, the number of sick days due to BP were also analysed. Since not every participant in the study held a daily sickness benefit insurance (in contrast to the statutory system), no monetary value was assigned to the days of sick leave.

Comparators

The intervention combined two care components: BP was treated by networks of general practitioners, orthopaedic surgeons and pain therapists working according to the FPZ concept and were located as close as possible to the patient's home [42]. This concept is based on a training programme for the spine-stabilising musculature developed at the German Sport University Cologne, which has been further advanced by the Cologne Research and Prevention Centre (FPZ) for the reconditioning of patients with sub-acute and chronic back pain (for details, see <http://www.fpz.de>). A therapy plan was drawn up following a functional biomechanical analysis comprised of up to 24 one-hour equipment-supported training units to build up the spine-stabilising musculature in FPZ back centres. After completing the training, IG members could receive €100 twice for the use of further freely selectable health sport offers as a movement bonus.

Each participant received personal telephone support from an external health coach provided by Thieme TeleCare. The aim of the coaching, which initially accompanied the therapy and was then followed up for six months, was to support behaviour changes, thus contributing to the continuation of physical activities.

Members of the CG did not receive any offer for training or coaching but were treated as current practice receiving standard care, for which recommendations exist in the *National Clinical Practice Guideline for Non-Specific Low Back Pain* [43].

Time horizon

The original study had a duration of 24 months [32]. The present cost comparison covered a period of four years. From the individual start time of the study (t-0), the costs were compared 24 months before and 24 months after the index date. Since Difference-in-Difference (DiD) estimation was chosen as the method of comparison (see methods section), the pre-intervention and post-intervention observation periods were selected to be equal.

The follow-up period of 24 months was chosen to compare the measured endpoint of health developments with the economic effects, thus calculating ICER using QALYs. No further tracking of cost developments into the future was carried out, as no longer-term data on the individual health status of the participants was collected.

Discount rate and currency

Following the recommendations of the German *Institute for Quality and Efficiency in Health Care* [IQWiG], the discount rate was set at 3 % per year [44]. Sensitivity analyses were performed for a discount rate of 0 % and 5 %.

Depending on which discount rate was applied, the results varied mildly. Since IQWiG recommends 3% and the overall results did not change much, only the results of 3 % discounting are presented in the analysis. Changes in statistical significance due to discounting are marked.

All costs were converted to 2017 Euros (€) using consumer price indices.

Choice of health outcomes

Outcomes of patients receiving standard care with those of a cohort receiving standard care combined with the intervention over two periods of 24 months were compared. The dataset provided by the insurance contained longitudinal patient-level information on medical diagnoses, direct medical costs (in 2017€) and healthcare utilisation between 2010 and 2017.

Values to calculate the QALYs, the number of sick days and the health status at the beginning of the study and after 24 months were obtained through an online questionnaire. The exact method is described elsewhere [32].

The study outcomes can be divided into direct medical costs (overall as well as BP specific), the number of sick days due to BP (*"Approximately how many days in the last six months were you unable to carry out your normal activities (work, school/study, housework) due to your back pain?"*) and the overall health developments (EQ-5D). BP related costs were identified through ICD-10 M40 to M54.9 diagnoses. All outcomes represent the average values over the 24-month baseline and follow-up period.

To calculate the QALYs, the EQ-5D was used. Since the original study used the SF-12, the EQ-5D value was calculated using Lawrence's algorithm [45].

Analytic methods

Study Size

A total number of 189 participants in the IG and 254 participants in the CG took part in the reference study [32]. After data cleaning and matching, 112 participants in the IG and 111 in the CG were included for further analysis. The selection criteria are shown in Table 2.

Participants who did not exercise physically due to a large distance to the training centre were excluded from the baseline (n=34) but controlled for in the second sensitivity analysis[1]. Participants who dropped out during a later stage (while training or during coaching) were treated according to the intention to treat principle and kept in the study.

Table 2 - Data preparation processes for selection of study population				
Data Processing Steps showing the number of participants	Overall	IG	CG	Used in
<i>Evaluation study group</i>	443	189	254	Publication in [32]
<i>Exclusion of participants without any billing invoice available</i>	435	186	249	Sensitivity Analysis III
<i>Exclusion due to unstable unit treatment</i>	431	185	246	
<i>Exclusion due to deductible/tariff</i>	404	172	232	
<i>Propensity score matching (Group ITT) + truncation</i>	273	136	137	Sensitivity Analysis II
<i>Exclusion of non-exercising participants</i>	223	112	111	Main Analysis
<i>Analysis of participants who improved their BP at the end of follow-up</i>	106	62	44	Sensitivity Analysis I

Table 2 - Data preparation process for selection of study population

Considering the stable unit treatment value assumption (STUVA) for the estimation of treatment effects [46], four further participants had to be excluded. One participant from the IG was excluded as he enrolled himself in the programme a second time before the end of the follow-up period. In the CG, three participants had to be excluded as they enrolled themselves in the intervention before the end of the study period.

As the data was provided by a PHI, there was the additional obstacle of managing the individual yearly threshold of costs before payment of expenses (deductible). To reduce the bias introduced by the tariff, 27 participants who did not hand in an invoice in one of the four examined years were excluded (yearly average of invoices = 42). Data management and statistical analysis were carried out using the software R 3.6.0 [47] in the application of the packages listed in the bibliography [48–54].

Difference-in-Difference regression

The aim of the DiD method is to estimate the average effect of treatment on the treated (ATT). Changes in cost over time between the IG and the CG were compared with a regression model with an interaction term between period and treatment ($Y = \beta_0 + \beta_1 * [\text{Period}] + \beta_2 * [\text{Treatment}] + \beta_3 * [\text{Period} * \text{Treatment}]$). The outcomes in the baseline period were measured two years before the respective index date. The time-series dimension of the two year baseline period was removed by comparing the values over two years to avoid biased standard errors due to serial correlation [55]. Thus, a single value per outcome measure for the baseline and follow-up period was generated.

To examine if the costs developed equally over time (parallel trend assumption) [56] BP specific costs were plotted over a time horizon of four years prior to the intervention. Outcomes were considered quarterly to test the parallel trend over 16 data points. The graph (Fig. 1) shows that the parallel trend assumption could be accepted. BP specific costs in both groups developed similarly before the start of the intervention. The average cost before the intervention could be divided by BP impairment.

Propensity score matching

Table 1 shows that IG and CG differed in the distribution of participants according to their pain levels. As the pain stage has been shown to have an impact on BP specific costs [5], nearest neighbour propensity score matching (PSM) with a caliper of 0.1 was performed to achieve adequate covariate balance at baseline [57]. Covariates used for the matching were sex, CCI, GCPS, STarT-Back, health cost before the intervention, age, sick days prior and risk for chronic BP as identified in the data selection [32].

As DiD regression is sensitive to high-leverage observations [58] extreme outliers (BP specific cost before intervention \geq €13.000) were excluded (n= 25). After excluding high-cost cases, the regression results became more robust.

Analysis of covariance (ANCOVA)

A one-sided covariance analysis (ANCOVA) was carried out for the outcomes of secondary interest here, i.e. the reduction of sick days due to BP and the general state of health (EQ-5D).

ANCOVA is a common, statistically robust method with model assumptions that should be respected (e.g. linearity between the covariate and the outcome, homogeneity of regression slopes, normally distributed outcome variable, homoscedasticity of residual variance) [59]. In the analysis of the EQ-5D, all model assumptions of the ANCOVA were met. For the range of sick days, however, the residuals were not normally distributed. This was, however, negligible as violations of these assumptions do not decisively influence either the probability of a first type error or the test strength [60, 61]. Covariance analyses are only contraindicated if the regression slopes are heterogeneous, the samples are unequal in size, and the residuals are not normally distributed [62]. This was not the case here. Therefore, an ANCOVA can be applied.

[1] Invitations were not controlled for the distance between the FPZ training centre and home. Some participants therefore only realised after their enrolment that the distance to reach the training centre was not manageable on a regular basis. The drop-out of these participants did not depend on their pain-levels as the percentage was comparable with 4.7% and 6.6 % dropout in the respective pain groups.

Results

Incremental costs and outcomes

Table 3 presents the cost data of the 223 participants analysed. Total medical costs were reduced in the IG and slightly increased in the CG, resulting in a statistically non-significant difference of - €769.16 (p = 0.78). BP specific costs were reduced in the IG and slightly increased in the CG so that an ATT of - €1096.56 (p = 0.039) was reached. The main driver for the reduction could be revealed as the reduced inpatient costs by - €717.00 (p = 0.025). The number of sick days due to BP in the last six months was reduced in the IG, whereas it increased in the CG. The estimated mean treatment difference was -17.5 days (p = 0.001) in the IG. The development in the EQ-5D was found to be more positive in the IG, compared to the CG (0.046, p= 0.026).

Table 3 - Discounted outcomes for the intervention (IG) and control group in the baseline (2 years) and follow-up period (2 years) with the respective Difference-in-Difference estimator and its standard error (SE)						
	IG (n =111)		CG (n=112)		DID Estimation	
Item	Baseline	Follow-up	Baseline	Follow-up	ATT	SE
<i>Total Cost (€)</i>	15699	14027	14887	13984	-769	2695
<i>Due to BP (€)</i>	3260	2095	2588	2519	-1097*	525
<i>Inpatient Cost – BP (€)</i>	869	242	403	494	-717*	315
<i>Outpatient Cost – BP (€)</i>	2391	1852	2185	2026	-380	404
<i>Sick Leave last six month due to BP^b</i>	31.3	18.44	31.3	36	-17.5**	3.57
<i>Overall Health Status (EQ-5D)^b</i>	0.604	0.678	0.604	0.632	0.046*	0.01
† < 0.1; * < 0.05; ** < 0.01; *** < 0.0001						
a: Average treatment effect for the treated represents excess resource utilisation attributable to DMP						
b: These results were not calculated with a DiD regression but with an ANCOVA						

Table 3 - Discounted outcomes for the intervention (IG) and control group (CG) in the baseline (2 years) and follow-up period (2 years) with the respective Difference-in-Difference estimator and its standard error (SE)

Subgroup analysis

In general BP specific costs differ depending on their GPCS status: The higher the status, the higher the costs [5]. To see if this applied to this data set, IG and CG participants were divided according to their BP impairment into *minor* and *major* impaired groups, as presented in Table 4.

In the *minor* impaired group, the total medical cost (€570.09, $p = 0.85$) and the back specific costs (-€538.10, $p = 0.42$) did not differ significantly. The amount of sick leave was significantly lower in the IG than in the CG (-9.49, $p = 0.024$). The overall health status was improved in both groups but did not reach statistical significance in between them (0.039, $p = 0.119$).

For the *major* impaired group, no significant difference in the overall direct cost could be revealed (-€2,107.90, $p = 0.62$). However, the ATT in BP specific cost was significant with a reduction by -€1,732.75 ($p = 0.035$) in favour of IG, which was mainly driven by a reduction in hospital costs (-€1,385.65, $p = 0.016$). The difference in sick leave in the last six months due to BP was 27 days ($p = 0.006$). The development of the overall health status showed an improvement by 0.047 points, albeit missing significance ($p = 0.158$).

Table 4 - Subgroup analysis according to impairment by back pain at baseline						
Minor Impaired (1)	IG (n =59)		CG (n=61)		DID	
Item	Baseline	Follow-up	Baseline	Follow-up	ATT	SE
<i>Total Cost (€)</i>	13804	14896	10578	11099	570	3252
<i>Due to BP (€)</i>	2768	2025	2051	1846	-538	661
<i>Inpatient Cost – BP (€)</i>	548	176	404	179	-147	323
<i>Outpatient Cost – BP (€)</i>	2219	1849	1646	1667	-391	539
<i>Sick Leave last six month due to BP^b</i>	3.3	5.8	3.3	15.3	-9.5*	2.91
<i>Overall Health Status (EQ-5D)^b</i>	0.699	0.749	0.699	0.710	0.039	0.02
Major Impaired (2)	IG (n=53)		CG (n=50)		DID	
Item	Baseline	Follow-up	Baseline	Follow-up	ATT	SE
<i>Total Cost (€)</i>	17809	13061	20145	17505	-2108	4297
<i>Due to BP (€)</i>	3808	2172	3244	3341	-1733*	813
<i>Inpatient Cost – BP (€)</i>	1226	317	401	878	-1386*	565
<i>Outpatient Cost –</i>	2582	1855	2843	2463	-347	597

BP (€)						
<i>Sick Leave last six month due to BP^b</i>	64	33.2	64	60.4	-27.2**	7
<i>Overall Health Status (EQ-5D)^b</i>	0.494	0.591	0.494	0.544	0.047	0.02

† < 0.1; * < 0.05; ** < 0.01; *** < 0.0001

a: Average treatment effect for the treated represents excess resource utilisation attributable to DMP

b: These results were not calculated with a DiD regression, but with an ANCOVA

Table 4 - Difference-in-Difference estimators (ATT) and their respective standard errors (SE) for two subgroup analyses: (1) Minor impairment by BP [GCPS grades I & II] and (2) Major impairment by BP [GCPS grades III & IV] at study enrolment status

ICER

The ICER was calculated using the discounted ATT and EQ-5D difference for the two years as stated above as well as the intervention cost of €1,500 which have not been included so far. To avoid the Keeler and Kretin paradoxon [63] costs and effects were equally discounted leading to the following equation:

$$\rightarrow \frac{(ATT \text{ Total Cost discounted}) + Interventioncost}{((Difference \text{ in } EQ-5D) * 2 \text{ years}) * Discount} = \frac{(-769.16) + 1500}{((0.046) * 2) * (\frac{1}{(1+0.03)^2})} = \text{€}8,427$$

According to the threshold of £20,000 to £30,000 per quality-adjusted life year set by the National Institute for Health and Care Excellence (NICE) [64] with an ICER of €8,427 per QALY gained, the intervention can be seen as cost-effective. If one would be able to steer the participants before enrolment better and assign them to the appropriate subgroup, one could achieve an ICER of -€6,861 per QALY. In doing so, one would create a dominant intervention that simultaneously saves costs and reveals a better outcome.

Calculating the ratio only with cost differences in BP specific cost reduces the ICER to €4,652 for the main group and -€2,627 for the subgroup with major impaired participants.

Characterising uncertainty

One source of uncertainty was the BP specific costs. Those were accounted for as soon as a diagnosis of the ICD group M40 to M54 was indicated on the bill. However, other diagnoses were often co-included on those invoices, resulting in ambivalence in cost allocation. In cases of doubt, it was not possible to distinguish precisely which costs were to be allocated to which diagnosis on the bill. This problem concerned members of IG and CG in equal shares and arose mainly in the outpatient sector. In order to compare whether cost-intensive co-diagnoses were more frequently coded in one group, the BP specific invoices were analysed additionally.

Table 5 - Characterising uncertainty of BP specific invoices				
	IG (n = 111)		CG (n = 112)	
Time	Baseline	Follow-up	Baseline	Follow-up
<i>Median</i>	10	9	8	8
<i>M40-54 Diagnosis alone</i>	33.2 %	27.4 %	31.1 %	27.6 %
<i>with other Diagnosis</i>	66.8 %	72.6 %	68.9 %	72.4 %
<i>Rank 1 ICD Group (%)</i>	M (25 %)	M (25 %)	M (21 %)	M (23 %)
<i>Rank 2 ICD Group (%)</i>	E (14%)	E (15 %)	E (11 %)	E (12 %)
<i>Rank 3 ICD Group (%)</i>	I (8%)	R (9 %)	R (10 %)	R (11%)
<i>Rank 4 ICD Group (%)</i>	G (8 %)	I (8%)	K (10 %)	K (10 %)
<i>Rank 5 ICD Group (%)</i>	R (8 %)	F (8 %)	I (9 %)	I (10 %)

Table 5 - Characterising uncertainty of BP specific invoices comparing median of BP invoices and distribution and frequency of co-listed diagnosis at baseline and follow-up period

The main comorbidities were the same in the first two groups and only differed from rank four onwards. In the group of other M diagnoses, gonarthrosis was mentioned most frequently, in the E group diabetes mellitus, in the I group mainly essential hypertension, in the R group mainly other chronic pain, in the K group fatty liver, in the G group polyneuropathy and in the F group an unspecified depressive episode. Table 5 shows how many invoices with M40-54 diagnoses were taken into account on average, how many invoices only contain an M40-54 diagnosis and the most frequently listed comorbidity groups. As can be seen in Table 5, the included bills are highly comparable to each other. It is noticeable that the IG had in median two more bills containing M diagnosis before the start of the programme. In the follow-up, however, this difference was diminished.

Due to minor differences in the distribution of the M diagnoses and the co-mentioned groups, it could be assumed that the BP specific costs in both groups were not influenced by a wrong allocation.

Characterising heterogeneity

The results of the main analysis indicated that participation in an MBR could save costs in the long run. The reduction in BP specific costs is significant both in the overall comparison and in the subgroup with *major* impairment. Due to the large standard deviation, the effect on total costs could not be clearly interpreted. In order to understand the development of the total costs better, a sensitivity analysis was carried out in which only participants indicating an improvement in BP - measured as a lower STarT-Back raw value accompanied by no regression in GCPS – were included ("profiteer" subgroup).

A second sensitivity analysis was conducted with all participants who enrolled in the programme, regardless of whether they finished all components or not (ITT). Explicitly included were those who could only take part in the coaching and not the physical training.

As matching in itself was – due to the small study group - a possible source of bias, a third sensitivity analysis compared the results without PSM.

Sensitivity Analysis I: Results of the outcome-based analysis

Table 6 presents the first sensitivity analysis with participants who improved their BP. In the IG, 62 of the 111 (56 %) participants achieved an improvement of their BP after two years. In the CG, only 44 of 112 (39 %) did. The total medical costs in the baseline period of the "profiteer" subgroup showed a range of €118,000. To exclude outliers, the groups were divided into low and high cost (0.1 – 0.5, 0.51 – 0.9 of the quantiles) members.

In the low-cost group, no significant difference in the costs between IG and CG could be identified. Only the improvement in the overall health status was different and in favour of the IG (0.11, $p = 0.034$).

In the high-cost group the total costs could be reduced by €5,431 ($p = 0.08$) in the IG compared to the CG (undiscounted: - €6,762, $p = 0.04$).

The number of sick days decreased significantly in the low and high-cost groups; to a greater extent in the IG than in the CG (low: -5.7, $p = 0.44$, high: - 18.3, $p = 0.06$). Statistical significance prevails below a level of 0.1 for the high-cost group.

Table 6 - Sensitivity analysis I with participants who improved their back pain						
Low Cost (0.1-0.5 Quantile)	IG (n =24)		CG (n=10)		DID	
Item	Baseline	Follow-up	Baseline	Follow-up	ATT	SE
<i>Total Cost (€)</i>	9159	9681	8274	6517	2280	1895
<i>Due to BP (€)</i>	1945	1273	1280	1347	-739	796
<i>Inpatient Cost – BP (€)</i>	359	67	0	194	-486	442
<i>Outpatient Cost – BP (€)</i>	1586	1205	1280	1152	-252	687
<i>Sick Leave last six month due to BP^b</i>	32.8	8.8	33.8	14.5	-5.7	6.04
<i>Overall Health Status (EQ-5D)^b</i>	0.607	0.745	0.607	0.634	0.111*	0.04
High Cost (0.5-0.9 Quantile)	IG (n=25)		CG (n=24)		DID	
Item	Baseline	Follow-up	Baseline	Follow-up	ATT	SE
<i>Total Cost (€)</i>	22172	12821	22069	18195	-5431†	3116
<i>Due to BP (€)</i>	7147	2913	5952	4210	-2492	1997
<i>Inpatient Cost – BP (€)</i>	2545	0	2695	184	-35	1395

<i>Outpatient Cost – BP (€)</i>	4601	2913	3257	4026	-2457†	1439
<i>Sick Leave last six month due to BP^b</i>	44.45	9.42	44.35	27.73	-18†	6.7
<i>Overall Health Status (EQ-5D)^b</i>	0.542	0.717	0.542	0.678	0.039	0.03
<p>† < 0.1; * < 0.05; ** < 0.01; *** < 0.0001</p> <p>^a: Average treatment effect for the treated represents excess resource utilisation attributable to DMP</p> <p>^b: These results were not calculated with a DiD regression but with an ANCOVA</p>						

Table 6 - Sensitivity analysis of participants who improved their BP with the respective Difference-in-Difference estimator (ATT) and its standard error (SE) in the baseline and follow-up period

Sensitivity Analysis II: Results Intention-to-treat

Table 7 shows the second sensitivity analysis with all participants who enrolled in the programme, regardless of whether they finished all components or not (ITT). Results of the second sensitivity analysis showed no significant result on the treatment effect on costs or health. The reduction of sick leave due to BP was significant. In the complete group, the IG had an ATT of 13.6 days (p= 0.002). The *major* impaired group (n= 62) reduced their sick leave due to BP by 22 days (p = 0.012) more than the CG (n = 57). In contrast to that, the *minor* impaired group (n= 74) reduced their sick leave due to BP by 7.2 days (p = 0.085) compared to the CG (n = 80).

Table 7 - Sensitivity analysis of participants who improved their BP with the respective Difference-in-Difference estimator (ATT) and its standard error (SE) in the baseline and follow-up period

Intention-to-treat (1)	All		Minor		Major	
	n: (IG = 136, CG = 137)		n: (IG = 74, CG = 80)		n: (IG = 62, CG = 57)	
Item	ATT	SE	ATT	SE	ATT	SE
<i>Total Cost (€)</i>	671	2310	2825	2714	-1782	3865
<i>Due to BP (€)</i>	-116	628	545	812	-907	977
<i>Inpatient Cost – BP (€)</i>	218	500	912	628	-660	811
<i>Outpatient Cost – BP (€)</i>	-334	354	-367	474	-248	528
<i>Sick Leave last six month due to BP^b</i>	-13.58	3.13	-7.24†	2.9	-22.04*	6.24
<i>Overall Health Status (EQ-5D)^b</i>	0.027	0.01	0.016	0.02	0.025	0.02
Original evaluation study (2)	All		Minor		Major	
	n: (IG = 186, CG = 249)		n: (IG = 88, CG = 147)		n: (IG = 98, CG = 102)	
Item	ATT	SE	ATT	SE	ATT	SE
<i>Total Cost (€)</i>	2594	2203	4982†	2695	987	3527
<i>Due to BP (€)</i>	490	811	955	929	430	1355
<i>Inpatient Cost – BP (€)</i>	601	651	1005	736	494	1105

Outpatient Cost – BP (€)	-111	354	-50	478	-64	529
Sick Leave last six month due to BP^b	-12.7**	2.53	-4.68	9.42	-22.84**	4.95
Overall Health Status (EQ-5D)^b	0.030*	0.01	0.011	0.01	0.048*	0.02

† < 0.1; * < 0.05; ** < 0.01; *** < 0.0001

a: Average treatment effect for the treated represents excess resource utilisation attributable to DMP

b: These results were not calculated with a DiD regression but with an ANCOVA

Table 7 - Sensitivity analysis: two scenarios (1: Intention-to-treat, 2: original evaluation study) with Difference-in-Difference estimators (ATT) and its standard error (SE)

Sensitivity Analysis III: Results without matching

Since the results and their significance level seem dependent on the matching, a third sensitivity analysis (second part of Table 7) was run with the original evaluation group [32]. The complete data preparation was omitted (no consideration of dropouts, tariff, or high-cost cases). Only those with no billing information over the course of four years (n=8) were excluded. In the overall population and the *major* impaired group, there was no significant cost difference between participants and non-participants in the programme. The *minor* impaired group has a favourable ATT of €4,982 (p = 0.065).

The health status and the number of sick days were in favour of the IG. In the whole group, the ATT for sick days is close to 13 days (p = 0.001), and the health gain is 0.03 (p = 0.039). In the *major* impaired group, participation in the health programme reduces sick days by almost 23 days (p = 0.001). The health status increases by 0.05 points (p = 0.030).

Discussion

Summary

The evaluation presented included real-world evidence from an outpatient, long-term MBR with behaviour-change coaching and device-supported exercise for chronic BP in the health care system of a PHI company in Germany between 2013 and 2017. It was demonstrated that the classification of the

participants according to their individual degree of impairment at the beginning - represented by the GCPS - is a very good separator for the effectiveness.

Significant cost savings were achieved especially in the *major* impaired group at the isolated BP specific costs (- €1,732). It is noteworthy that participants in the IG needed at least one additional doctoral visit before the start of the training because a medical prescription for the FPZ training was required, often combined with an MRI to rule out medical contraindication. Despite these extra charges, the IG still reduced BP related costs significantly, indicating that the intervention is saving costs in the longer term. This is in line with the results of Müller et al. [27]: In a multimodal intervention with a similar exercise therapy in Germany's SHI, the savings depended on the BP impairment at baseline. Their intervention was particularly cost-saving for participants with GCPS grades IV. In contrast to Müller's study, the intervention examined here can be classified as an MBR with fewer training units but with an additional coaching component also focusing on behavioural change. We conclude that BP specific cost savings can be already achieved from GCPS grade III onwards.

With a QALY gain of 0.046, the intervention achieved the second-best benefit compared to all forms of therapy presented by Herman et al. [15]. The MBR duration of one year resulted in high programme costs of €1,500. Savings of €769 in total costs, respectively €1,097 in BP specific costs resulted in an ICER of €8,428 (respectively €4,652 for BP specific costs only) per QALY gained. The intervention considered in this study can therefore be classified as cost-effective.

If the insurance companies were able to steer the participants better before enrolment and assign them in *minor* or *major* impaired appropriately, a dominant intervention with an ICER of - €6,861 could be achieved. Herman et al. have shown that most interventions for CLBP are cost-effective from the perspective of the payer - whereas they are dominant for the society [15]. Focus on the direct medical costs results in cost-effectiveness, but the additional consideration of the days of sick leave saved (-17.5 in the last six month) turns the intervention into a cost-saving instrument from society's perspective. Furthermore, it could be shown that by consistently using the GCPS to determine participation in the programme, it is possible to create a dominant intervention from the payer's point of view.

The study was able to observe that BP prevailed in 61% of the participants in the CG after two years, which is subject to standard care. This correlates with BP being considered a long-term condition with a variable course [65]. The exact recurrence is unclear, but 33% to 67% of people with BP can be expected to have permanent recurrent episodes [6]. The present study adds to the existing knowledge that if BP is improved, the outcome differs strongly in terms of cost. The total costs are reduced in both "profiteer" groups. However, the IG has more savings in total costs than the CG (- €5,431). By focusing on exercise and self-efficacy, the high total cost can be significantly reduced. Judging from a pure cost perspective, it can be stated that the payer should increase its efforts to persuade the severely affected and cost-intensive insureds to participate in an MBR – and find an alternative offer for the *minor* impaired & low-cost groups.

However, if the perspective of society is also taken into account, participation in a back programme is worthwhile for all those affected. Across all groups and sensitivity analyses, the reduction in sick days due to BP was significantly higher in the IG than in the CG (ranging from 9.5 to 27 days). Wagner et al. reported a reduction of 44.3 days in the duration of incapacity to work and BP specific saving of €1,284 (daily sickness allowance excluded) after the completion of a 20 days short-term IMPT [13]. The effects were observed after one year - however, their programme is more intensive (time and costs) with more than 100 hours of treatment and twice the expenses. The target group is comparable to the *major* group, and the results are similar, even if the period under consideration is different. If one takes into account that there was a reduction of 27 sick days in the last six months, it can be assumed that a long term MBR is not inferior to a focused, more expensive IMPT in terms of sick days and cost reduction. Considering the results of the *major* impaired group – this intervention can become superior in access, cost-savings and reduction in sick leave - if the provider finds a way to allocate groups more targeted.

Sensitivity analyses II and III indicate that the payer should also aim to ensure that the MBRs created are carried out in full and analysed thoroughly with regard to cost effects. Special cost background of the privately insured, with the possibility of not submitting incurred costs on the one hand, and the heterogeneity of the population with regard to the overall health burden on the other need to be taken into account. This enforces the need to conduct the cost analysis with a large population that can be better and more clearly separated.

The sensitivity analyses nevertheless confirm the positive effect of the intervention on sick days and general health, which suggests that these effects are robust.

A DMP for BP in Germany is long overdue. It has been proven that significant improvements in health and savings of direct and societal costs are possible through improved management of CLBP. With an excellent QALY gain, cost savings in the treatment of BP and high reductions in sick days, the present cost-effective MBR can serve as a blueprint for a treatment programme that applies the clinical guidelines and is both medically and economically effective.

Limitations

The study has three limitations: BP specific costs, size and the time perspective.

The first limitation is the calculation of the BP specific costs. By analysing the comorbidities in detail, an attempt was made to keep the influence of other diseases as low as possible. A residual doubt remains that could only be remedied in the setting by focusing solely on inpatient visits due to BP. Since a large proportion of the costs occur in the outpatient sector and only a small number of the study population was treated as inpatients, this was - in the interest group size - not done here. With a larger group, only taking inpatient costs could be a feasible approach in the clarification of BP specific costs. It should also be noted that higher total costs are to be expected in the SHI system. In contrast to the PHI, sickness benefits must be paid after a certain period of absence. Therefore, it can be assumed that the cost-benefit effect of a long-term MBR treatment would be better in the SHI.

Even though the sample size with 443 participants was more extensive than most other published studies [9], it was still challenging to present clear results – making this the second limitation. The sample size was reduced by the exclusion of dropouts, switchers, tariff-related exclusion, truncation and PSM. In the course of the analysis, the still cost heterogeneous groups needed to be separated according to BP status. The analysis suggests savings from participation in the MBR in many places – but occasionally does not meet significance levels. Thus, those findings should be verified in a study with a larger sample size.

The third limitation is the time perspective. A study period of four years exceeds other published analyses but might still exclude savings that occur further down the line. Generali's business case is calculated for five years after the start of the programme, which suggests that more savings are expected later on. For a complex intervention like the one in the present study, the *Medical Research Council* recommends a lifetime horizon to demonstrate the sustainability of outcome changes [66]. Since no information was available on the self-assessed health status after the follow-up period, this could not be done in the present study. However, this is the first health cost-utility study with data from a PHI in Germany on a health intervention that is actually offered, run and followed up thoroughly and accompanied by a trial study. The presented results therefore are able to address the aforementioned uncertainty gap about the cost-effectiveness of long-term MBR.

Generalisability

The results can be generalised to a limited extent. The data used was granted by a PHI provider, who is, in general, insuring healthier individuals than an SHI [67]. It can be assumed that the high improvement rate in BP in the CG is due to the composition of health insurance risk profiles. Good insurance risks are known to take care of their own health and believe in self-efficacy [68, 69]. The effect in favour of the IG might be higher in the general population, where a more passive CG can be expected.

With a gender ratio of 65(M) to 35(F), there is a male surplus, which is not representative of the overall German population but can be considered representative for PHI [70]. To achieve better comparability with the statutory system, specific individual costs (e.g. one/two-bedroom) were not considered. Additionally, the individual's tariff influence was kept as low as possible by taking the billed amount as cost and not the respective amount of refund.

The perspective of the analysis was stated from the payer's point of view; indirect medical costs incurred by the individual, their relatives, society or the employer were not taken into account. The apparent savings in sick days and the improvement in general health status suggest that further savings beyond direct medical costs are conceivable.

The data used in the analysis was obtained from a single clinical trial, which potentially limits the generalisability of the findings for several reasons [71]. However, since the study was multicentric and the participants and study centres were spread across Germany, this factor was mitigated.

The effects shown in this study using a relatively small sample size suggest that a participation in an outpatient MBR with behaviour-change coaching and device-supported exercise leads to cost savings, reduction of sick days as well as improvements in health status. The different effects were backed up with extensive sensitivity analyses, which increased the robustness of the findings. To erase the still prevailing doubt, an evaluation with a larger group is recommended.

Conclusion

This is the first cost-utility study with combined data from a PHI and a controlled trial in Germany to demonstrate that long term MBR is a cost-effective instrument in the treatment of CLBP at 24-month follow-up. Subgroups with major impairment measured by Korff's Chronic Pain Grade Questionnaire benefit more from the intervention than patients with minor impairment due to BP. When the findings are compared to more expensive IMPT in terms of sick days and cost reduction - the intervention analysed here can become superior in access, cost savings and reduction in sick leave. It is therefore recommended to use MBR for the major impairment group. For the minor impaired another solution needs to be developed. Providers should try to identify BP impairment in their administrative data so that they can make targeted offers. It should be stressed that MBR significantly reduces sick leave in minor and major impaired participants in the long term, making it a profitable intervention from society's point of view. However, it remains to be seen how those positive results develop over a lifetime horizon.

Declarations

Funding: In this study, data from the published medical evaluation are linked with economic administrative data to carry out a cost-benefit analysis. The medical evaluation study was funded by the Generali Deutschland Krankenversicherung AG. It paid a grant for the scientific evaluation of the programme to the researchers of the University of Lübeck.

Acknowledgements: We would like to thank Generali Deutschland Krankenversicherung AG and the University of Lübeck for providing access and support in obtaining the data.

Competing interests: Martin Hochheim (MH) is working part time for the Generali Health Solutions GmbH (GHS), which is affiliated with the Generali Deutschland Krankenversicherung. Max Wunderlich (MW) is managing director of the GHS. Philipp Ramm (PR) is currently responsible for the BP programme. They declare that research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. Volker Amelung (VA) declares no conflicts of interest.

Availability of data and material (data transparency): The data that support the findings of this study are available from Generali Deutschland Krankenversicherung AG and University of Lübeck but restrictions apply to the availability of these data, which were used under license for the current study, and so are not

publicly available. The data is however available from the authors upon reasonable request and with permission including a signed data access agreement of Generali Deutschland Krankenversicherung AG and University of Lübeck.

Code availability: Data management and statistical analysis were carried out using the software R in application of the packages listed in the bibliography. The code that supports the findings of this study is available from Generali Deutschland Krankenversicherung AG but restrictions apply. It is available from the authors upon reasonable request and with permission including a signed access agreement of Generali Deutschland Krankenversicherung AG.

Authors' contributions: MH planned the analyses, analysed the data, interpreted the results, wrote the manuscript. VA supervised the project. MW contributed to the implementation of the research. JPS verified the analytical methods. PR contributed to the interpretation of the results. All authors provided critical feedback and helped shape the research, analysis and manuscript. All authors read and approved the final manuscript.

Trial registration: The trial of the evaluation study was registered at the German Clinical Trials Register under DRKS00015463 retrospectively (dated 4 Sept 2018).

Ethics approval and consent to participate: The independent research ethics committee of the University of Lübeck gave approval for the medical evaluation study (Re.-No.14 –249, dated 20.11.2014). As the participants had already consented to the usage of the data for further analysis, no new ethic vote was sought for the present evaluation. Written informed consent was obtained from all study participants.

Consent for publication: Not applicable

References

1. James SL, Abate D, Abate KH, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet*. 2018;392:1789–858. [https://doi.org/10.1016/S0140-6736\(18\)32279-7](https://doi.org/10.1016/S0140-6736(18)32279-7).
2. Maher C, Underwood M, Buchbinder R. Non-specific low back pain. *Lancet*. 2017;389:736–47. [https://doi.org/10.1016/S0140-6736\(16\)30970-9](https://doi.org/10.1016/S0140-6736(16)30970-9).
3. Walker BF, Muller R, Grant WD. Low back pain in Australian adults: The economic burden. *Asia-Pacific J Public Heal*. 2003;15:79–87. <https://doi.org/10.1177/101053950301500202>.
4. Luo X, Pietrobon R, Sun SX, Liu GG, Hey L. Estimates and Patterns of Direct Health Care Expenditures among Individuals with Back Pain in the United States. *Spine (Phila Pa 1976)*. 2004;29:79–86. <https://doi.org/10.1097/01.BRS.0000105527.13866.0F>.
5. Wenig CM, Schmidt CO, Kohlmann T, Schweikert B. Costs of back pain in Germany. *Eur J Pain*. 2009;13:280–6. <https://doi.org/10.1016/j.ejpain.2008.04.005>.

6. Hartvigsen J, Hancock MJ, Kongsted A, et al. What low back pain is and why we need to pay attention. *Lancet*. 2018;391:2356–67. [https://doi.org/10.1016/S0140-6736\(18\)30480-X](https://doi.org/10.1016/S0140-6736(18)30480-X).
7. Kamper SJ, Apeldoorn AT, Chiarotto A, Smeets RJE, Ostelo RWJG, Guzman J, van Tulder MW. Multidisciplinary biopsychosocial rehabilitation for chronic low back pain: Cochrane systematic review and meta-analysis. *BMJ*. 2015;350:h444. <https://doi.org/10.1136/bmj.h444>.
8. Foster NE, Anema JR, Cherkin D, et al. Prevention and treatment of low back pain: evidence, challenges, and promising directions. *Lancet*. 2018;391:2368–83. [https://doi.org/10.1016/S0140-6736\(18\)30489-6](https://doi.org/10.1016/S0140-6736(18)30489-6).
9. Almazrou SH, Elliott RA, Knaggs RD, Alaujan SS. Cost-effectiveness of pain management services for chronic low back pain: A systematic review of published studies. *BMC Health Serv Res* 20, (2020). <https://doi.org/10.1186/s12913-020-5013-1>.
10. Costa LDCM, Koes BW, Pransky G, Borkan J, Maher CG, Smeets RJE. Primary care research priorities in low back pain: an update. *Spine (Phila Pa 1976)*. 2013;38:148–56. <https://doi.org/10.1097/BRS.0b013e318267a92f>.
11. Brömme J, Mohokum M, Disch AC, Marnitz U. Interdisziplinäre, multimodale Schmerztherapie vs. konventionelle Therapie: Eine Kostenanalyse bei Patienten mit chronischen Rückenschmerzen. *Schmerz*. 2015;29:195–202. <https://doi.org/10.1007/s00482-014-1508-1>.
12. Nagel B, Korb J. Multimodale Therapie. *Orthopade*. 2009;38:907–12. <https://doi.org/10.1007/s00132-009-1482-y>.
13. Wagner CJ, Ayyad G, Otzdorff A, Bienek K, Marnitz U, Pickardt B von, Seidel W, Sehlen S, Supantia P, Lindena G. Behandlungs- und Kosteneffekte der interdisziplinären multimodalen Schmerztherapie bei Patienten mit Rückenschmerz. *Der Schmerz*. 2019;33:128–38. <https://doi.org/10.1007/s00482-019-0356-4>.
14. Dragioti E, Björk M, Larsson B, Gerdle B. A Meta-Epidemiological Appraisal of the Effects of Interdisciplinary Multimodal Pain Therapy Dosing for Chronic Low Back Pain. *J Clin Med*. 2019;8:871. <https://doi.org/10.3390/jcm8060871>.
15. Herman PM, Lavelle TA, Sorbero ME, Hurwitz EL, Coulter ID: Are Nonpharmacologic Interventions for Chronic Low Back Pain More Cost Effective Than Usual Care? Proof of Concept Results from a Markov Model. *Spine (Phila. Pa. 1976)*. 44, 1456–1464 (2019). <https://doi.org/10.1097/BRS.0000000000003097>.
16. Herman PM, McBain RK, Broten N, Coulter ID: Update of Markov Model on the Cost-effectiveness of Nonpharmacologic Interventions for Chronic Low Back Pain Compared to Usual Care. *Spine (Phila. Pa. 1976)*. 45, 1383–1385 (2020). <https://doi.org/10.1097/BRS.0000000000003539>.
17. Lamb SE, Hansen Z, Lall R, Castelnovo E, Withers EJ, Nichols V, Potter R, Underwood MR. Group cognitive behavioural treatment for low-back pain in primary care: a randomised controlled trial and cost-effectiveness analysis. *Lancet*. 2010;375:916–23. [https://doi.org/10.1016/S0140-6736\(09\)62164-4](https://doi.org/10.1016/S0140-6736(09)62164-4).

18. Cambron JA, Gudavalli MR, Hedeker D, McGregor M, Jedlicka J, Keenum M, Ghanayem AJ, Patwardhan AG, Furner SE. One-year follow-up of a randomized clinical trial comparing flexion distraction with an exercise program for chronic low-back pain. *J Altern Complement Med*. 2006;12:659–68. <https://doi.org/10.1089/acm.2006.12.659>.
19. Sherman KJ, Cherkin DC, Erro J, Miglioretti DL, Deyo RA. Comparing yoga, exercise, and a self-care book for chronic low back pain: A randomized, controlled trial. *Ann Intern Med*. 2005;143:849–56. <https://doi.org/10.7326/0003-4819-143-12-200512200-00003>.
20. Cherkin DC, Sherman KJ, Avins AL, et al. A randomized trial comparing acupuncture, simulated acupuncture, and usual care for chronic low back pain. *Arch Intern Med*. 2009;169:858–66. <https://doi.org/10.1001/archinternmed.2009.65>.
21. Haas M, Vavrek D, Peterson D, Polissar N, Neradilek MB. Dose-response and efficacy of spinal manipulation for care of chronic low back pain: A randomized controlled trial. *Spine J*. 2014;14:1106–16. <https://doi.org/10.1016/j.spinee.2013.07.468>.
22. Martini L, Hoffmann F. Comorbidity of chronic back pain and depression in Germany: Results from the GEDA study, 2009 and 2010. *Z Evid Fortbild Qual Gesundheitswes*. 2018;137–138:62–8. <https://doi.org/10.1016/j.zefq.2018.10.003>.
23. Manchikanti L, Singh V, Falco FJE, Benyamin RM, Hirsch JA. Epidemiology of low back pain in Adults. *Neuromodulation*. 2014;17:3–10. <https://doi.org/10.1111/ner.12018>.
24. Groessl EJ, Liu L, Richard EL, Tally SR. Cost-effectiveness of Yoga for Chronic Low Back Pain in Veterans. *Med Care*. 2020;58:142–8. <https://doi.org/10.1097/MLR.0000000000001356>.
25. Lambeek LC, Bosmans JE, Van Royen BJ, Van Tulder MW, Van Mechelen W, Anema JR. Effect of integrated care for sick listed patients with chronic low back pain: economic evaluation alongside a randomised controlled trial. *BMJ* 341, (2010). <https://doi.org/10.1136/bmj.c6414>.
26. Williams A, van Dongen JM, Kamper SJ, et al. Economic evaluation of a healthy lifestyle intervention for chronic low back pain: A randomized controlled trial. *Eur J Pain*. 2019;23:621–34. <https://doi.org/10.1002/ejp.1334>.
27. Müller G, Pfänder M, Clement M, et al. Therapeutic and economic effects of multimodal back exercise: A controlled multicentre study. *J Rehabil Med*. 2019;51:61–70. <https://doi.org/10.2340/16501977-2497>.
28. Andronis L, Kinghorn P, Qiao S, Whitehurst DGT, Durrell S, McLeod H. Cost-Effectiveness of Non-Invasive and Non-Pharmacological Interventions for Low Back Pain: a Systematic Literature Review. *Appl Health Econ Health Policy*. 2017;15:173–201. <https://doi.org/10.1007/s40258-016-0268-8>.
29. Fatoye F, Wright JM, Gebrye T. Cost-effectiveness of physiotherapeutic interventions for low back pain: a systematic review. *Physiotherapy*. 2020;108:98–107. <https://doi.org/10.1016/j.physio.2020.04.010>.
30. Miyamoto GC, Lin C-WC, Cabral CMN, Dongen JM van, Tulder MW van. Cost-effectiveness of exercise therapy in the treatment of non-specific neck pain and low back pain: a systematic review with meta-analysis. *Br J Sports Med*. 2019;53:172–81. <https://doi.org/10.1136/BJSPORTS-2017-098765>.

31. Hüppe A, Wunderlich M, Hochheim M, Mirbach A, Zeuner C, Raspe H. Evaluation of a Proactive Health Programme for Insured Persons with Persistent Back Pain: One-year Follow-up of a Randomised Controlled Trial. *Gesundheitswesen*. 2017;81:831–8. <https://doi.org/10.1055/s-0043-121696>.
32. Hüppe A, Zeuner C, Karstens S, Hochheim M, Wunderlich M, Raspe H. Feasibility and long-term efficacy of a proactive health program in the treatment of chronic back pain: a randomized controlled trial. *BMC Health Serv Res*. 2019;19:714. <https://doi.org/10.1186/s12913-019-4561-8>.
33. Von Korff M, Ormel J, Keefe FJ, Dworkin SF. Grading the severity of chronic pain. *Pain*. 1992;50:133–49. [https://doi.org/10.1016/0304-3959\(92\)90154-4](https://doi.org/10.1016/0304-3959(92)90154-4).
34. Klasen BW, Hallner D, Schaub C, Willburger R, Hasenbring M. Validation and reliability of the German version of the Chronic Pain Grade questionnaire in primary care back pain patients. *Psychosoc Med*. 2004;1:Doc07.
35. Hill JC, Dunn KM, Lewis M, Mullis R, Main CJ, Foster NE, Hay EM. A primary care back pain screening tool: Identifying patient subgroups for initial treatment. *Arthritis Rheum*. 2008;59:632–41. <https://doi.org/10.1002/art.23563>.
36. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. *J Chronic Dis*. 1987;40:373–83. [https://doi.org/10.1016/0021-9681\(87\)90171-8](https://doi.org/10.1016/0021-9681(87)90171-8).
37. Hofmann A, Browne M. One-sided commitment in dynamic insurance contracts: Evidence from private health insurance in Germany. *J Risk Uncertain*. 2013;46:81–112. <https://doi.org/10.1007/s11166-012-9160-6>.
38. Institute for Quality and Efficiency in Health Care (IQWiG). Systematic guideline search and appraisal, as well as extraction of relevant recommendations, for a DMP “chronic back pain,” <https://www.iqwig.de/en/projects/v14-04.html>. Accessed April 5, 2021.
39. Amelung VE, Lägel R. *Innovative Konzepte und Verträge in der Rückenschmerztherapie*. Berlin: MWV Medizinisch Wissenschaftliche Verlagsgesellschaft; 2014.
40. Porter ME, Guth C. *Chancen für das deutsche Gesundheitssystem*. Berlin Heidelberg: Springer; 2012.
41. Central Krankenversicherung AG. *Geschäftsbericht 2019 Central Krankenversicherung AG*, <https://www.general.de/berichte/>. Accessed December 8, 2020.
42. Denner A. *Analyse und Training der wirbelsäulenstabilisierenden Muskulatur*. Berlin Heidelberg: Springer; 1998.
43. Oliveira CB, Maher CG, Pinto RZ, Traeger AC, Lin C-WC, Chenot J-F, van Tulder M, Koes BW. Clinical practice guidelines for the management of non-specific low back pain in primary care: an updated overview. *Eur Spine J*. 2018;27:2791–803. <https://doi.org/10.1007/s00586-018-5673-2>.
44. IQWiG. *Methoden A*. <https://www.iqwig.de/en/methods/methods-paper.3020.html>. Accessed December 21, 2020.
45. Lawrence WF, Fleishman JA. Predicting EuroQoL EQ-5D Preference Scores from the SF-12 Health Survey in a Nationally Representative Sample. *Med Decis Mak*. 2004;24:160–9.

- <https://doi.org/10.1177/0272989X04264015>.
46. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983;70:41–55. <https://doi.org/10.1093/biomet/70.1.41>.
 47. R Core Team: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. (2020). <https://www.R-project.org>. Accessed May 19, 2021.
 48. Wickham H, Averick M, Bryan J, et al. Welcome to the Tidyverse. *J Open Source Softw*. 2019;4:1686. <https://doi.org/10.21105/joss.01686>.
 49. Grolemund G, Wickham H. Dates and Times Made Easy with lubridate. *J Stat Softw*. 2011;40:1–25. <https://doi.org/10.18637/jss.v040.i03>.
 50. Yoshida K, Bartel A: tableone: Create “Table 1” to Describe Baseline Characteristics with or without Propensity Score Weights. R package version 0.12.0. (2020). <https://cran.r-project.org/package=tableone>. Accessed May 19, 2021.
 51. Ho DE, Imai K, King G, Stuart EA. MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. *J Stat Softw*. 2011;42:1–28. <https://doi.org/10.18637/jss.v042.i08>.
 52. Gasparini A. comorbidity: An R package for computing comorbidity scores. *J Open Source Softw*. 2018;3:648. <https://doi.org/10.21105/joss.00648>.
 53. Lenth RV: emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.5.5-1. (2021). <https://cran.r-project.org/package=emmeans>. Accessed May 19, 2021.
 54. Greifer N: cobalt: Covariate Balance Tables and Plots. R package version 4.3.1. (2021). <https://cran.r-project.org/package=cobalt%7D>. Accessed May 19, 2021.
 55. Bertrand M, Duflo E, Mullainathan S. How Much Should We Trust Differences-In-Differences Estimates? *Q. J Econ*. 2004;119:249–75. <https://doi.org/10.1162/003355304772839588>.
 56. Abadie A, Cattaneo MD. Econometric Methods for Program Evaluation. *Annu Rev Econom*. 2018;10:465–503. <https://doi.org/10.1146/annurev-economics-080217-053402>.
 57. Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behav Res*. 2011;46:399–424. <https://doi.org/10.1080/00273171.2011.568786>.
 58. Han JS, Houde J-F, van Benthem AA, Abito JM. When Does Regulation Distort Costs? Lessons from Fuel Procurement in US Electricity Generation: Comment. *Am. Econ Rev*. 2021;111:1356–72. <https://doi.org/10.1257/aer.20200679>.
 59. Pocock SJ, Assmann SE, Enos LE, Kasten LE. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. *Stat Med*. 2002;21:2917–30. <https://doi.org/10.1002/sim.1296>.
 60. Bortz J, Schuster C. Statistik für Human-und Sozialwissenschaftler. Lehrbuch mit Online-Materialien. Springer DE; 2010.
 61. Wang B, Ogburn EL, Rosenblum M. Analysis of Covariance (ANCOVA) in Randomized Trials: More Precision and Valid Confidence Intervals. Without Model Assumptions *Biometrics*. 2019;75:1391–

400. <https://doi.org/10.1111/biom.13062>.
62. Levy KJ. A Monte Carlo Study of Analysis of Covariance Under Violations of the Assumptions of Normality and Equal Regression Slopes. *Educ Psychol Meas.* 1980;40:835–40. <https://doi.org/10.1177/001316448004000404>.
63. Attema AE, Brouwer WBF, Claxton K. Discounting in Economic Evaluations. *Pharmacoeconomics.* 2018;36:745–58. <https://doi.org/10.1007/s40273-018-0672-z>.
64. National Institute for Health and Clinical Excellence (NICE). Guide to the methods of technology appraisal 2013, <https://www.nice.org.uk/process/pmg9/resources/guide-to-the-methods-of-technology-appraisal-2013-pdf-2007975843781>. Accessed January 29, 2021.
65. Dunn KM, Hestbaek L, Cassidy JD. Low back pain across the life course. *Best Pract Res Clin Rheumatol.* 2013;27:591–600. <https://doi.org/10.1016/j.berh.2013.09.007>.
66. Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, Moore L, O’Cathain A, Tinati T, Wight D, Baird J. Process evaluation of complex interventions: Medical Research Council guidance. *BMJ* 350, (2015). <https://doi.org/10.1136/bmj.h1258>.
67. Greß S. Private Health Insurance in Germany: Consequences of a Dual System. *Healthc Policy | Polit Santé.* 2007;3:29–37. <https://doi.org/10.12927/hcpol.2007.19389>.
68. de M Vieira, de Góes Salvetti ÉB, Damiani M, de Mattos Pimenta LP. C.A.: Self-Efficacy and Fear Avoidance Beliefs in Chronic Low Back Pain Patients: Coexistence and Associated Factors. *Pain Manag Nurs.* 2014;15:593–602. <https://doi.org/10.1016/j.pmn.2013.04.004>.
69. Lütke-Bornefeld S. Risikoselektion in der Krankenversicherung, <https://www.bertelsmann-stiftung.de/de/publikationen/publikation/did/spotlight-gesundheit-risikoselektion-in-der-krankenversicherung>. Accessed January 29, 2021.
70. Verband der Privaten Krankenversicherung e. V: Zahlenbericht. 2019, <https://www.pkv.de/verband/presse/meldungen-2020/pkv-verband-veroeffentlicht-zahlenbericht-2019>. Accessed January 20, 2021.
71. Stuart EA, Bradshaw CP, Leaf PJ. Assessing the Generalizability of Randomized Trial Results to Target Populations. *Prev Sci.* 2015;16:475–85. <https://doi.org/10.1007/s11121-014-0513-z>.

Figures

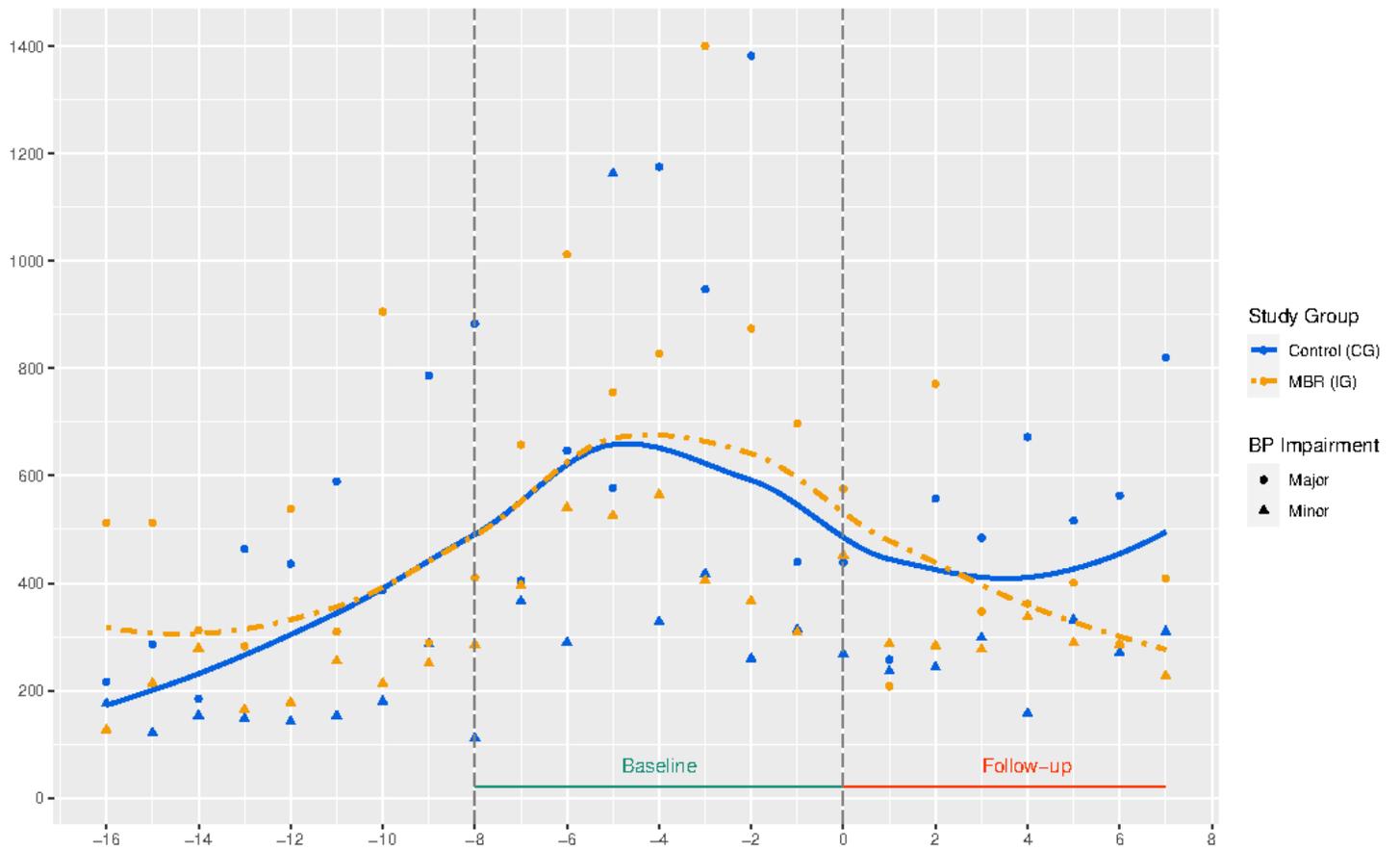


Figure 1

Parallel Trend Assumption showing the development of BP specific costs grouped by BP impairment and study group