Validation of a Back Pain Severity Prediction Algorithm: A Cross-Sectional Study with Updated Healthcare Costs for Back Pain Patients Based on the Graded Chronic Pain Scale

Martin Hochheim (✉ Martin.J.Hochheim@stud.mh-hannover.de)

Hannover Medical School

Philipp Ramm

Generali Health Solutions GmbH

Max Wunderlich

Generali Health Solutions GmbH

Volker Amelung

Hannover Medical School

Research Article

Keywords: Back Pain, Chronic Pain, Korff's Graded Chronic Pain Scale (GCPS), Cost of illness, Direct costs, Prediction, Claims data, Validity

Posted Date: August 2nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-699141/v1>

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

[Read Full License](#)

Abstract

Background: Treatment of chronic lower back pain (CLBP) should be stratified for best medical and economic outcome. To improve targeting of potential participants for exclusive therapy offers by payers, Freytag et al. developed an algorithm to identify back pain chronicity classes (CC) based on claims data. The aim of this study was the external validation of the algorithm, as this was previously lacking.

Methods: Administrative claims data and self-reported patient information of 3,506 participants of a health management programme of a private health insurance in Germany were used to validate the algorithm. Sensitivity, specificity and Matthews correlation coefficient (MCC) were computed comparing the prediction with actual grades based on von Korff's Graded Chronic Pain Scale (GCPS). Secondary outcome was an updated view on direct health care costs (€) of back pain (BP) patients grouped by GCPS.

Results: Results showed a *fair* correlation between predicted CC and actual GCPS grades. A total of 69.7 % of all cases were classified correctly. Sensitivity and specificity rates of 54.6 % and 76.4 % underlined the accuracy. Correlation between CC and GCPS with an MCC of 0.304 also indicated a *fair* relationship between prediction and observation. Cost data could be clearly grouped by GCPS: the higher the grade, the higher the costs and health care usage.

Conclusions: This was the first study to compare the predicted BP severity using claims data with actual BP severity by GCPS. Based on the results, the usage of the CC as a single tool to determine who receives treatment of CLBP cannot be recommended. The CC is a good tool to segment candidates for BP specific types of intervention. However, it cannot replace a medical screening at the beginning of an intervention as the rate of false negatives is too high.

Trial registration: The study was conducted using routinely collected data from an intervention, which was evaluated and registered previously at the German Clinical Trials Register under DRKS00015463 retrospectively (4 Sept 2018). The informed consent and the self-reported questionnaire have remained unchanged since the study and are therefore still valid in accordance with the ethics proposal.

Background

Low back pain (LBP) is a prevalent symptom occurring in all age groups. It is currently ranked as the most common reason for disability globally [1]. A prevalence of 7.5 % in 2017 indicates that 577 million people worldwide have experienced LBP in that year. The lifetime prevalence of LBP is as high as 70 % to 80 %, i.e. almost everyone will suffer from back pain (BP) during their lifetime [2–4]. Usually, an exact pathologic cause of LBP cannot be identified, which is why most cases are classified as non-specific [2, 5, 6].

LBP is often divided into three stages of chronicity based on the duration of an episode of pain: Acute (< 6 weeks), subacute (6–12 weeks) and chronic (> 12 weeks). The majority of those affected recover within

the first six weeks. However, 10 % to 40 % remain in a state of pain [3, 4]. But duration alone is not the only criterion for classifying chronic BP. Other influencing factors include the severity of pain, accompanying depressive symptoms with their corresponding severity, disability and recovery expectations among others [7–9].

Von Korff et al. have developed a model to assess the severity of chronic pain considering the duration, intensity and impairment of BP within the last six months. Its application leads to an individual classification into one of the 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*. Grades II to IV are clinically significant containing patients affected with intense BP with mild to severe dysfunction [7, 10]. It can be stated that the more severe the chronicity, the higher the limitation and the direct and indirect medical costs incurred. A sharp increase in costs in higher GCPs grades was observed elsewhere [11, 12].

Treatment of BP depends on the chronicity and focuses on different areas. After a short screening for “red flags”, which potentially indicates a specific cause and a more serious aetiology, patients with nonspecific acute BP are usually educated and encouraged to remain active. Due to the natural course of BP early interventions to prevent progression from acute to chronic state are unnecessary for the majority of patients [8, 13].

Structured physical and psychological interventions, as well as education, are recommended for patients whose BP is subacute or chronic and those with current acute pain, but at high risk of becoming chronic BP patients [4, 14, 15]. There are various approaches that are promising a favourable outcome [13, 16–19]. These interventions are effective, but also costly and limited in number [19, 20]. In addition, BP is often not treated according to guidelines and best practices. Access to and transitions from primary care to these interventions are still rare. Furthermore, there is too frequent and rapid use of imaging, spinal injection therapies and surgical procedures present in the treatment of CLBP [14]. In a recently published study, Daniel and colleagues used administrative data from a large statutory health insurance (SHI) in Germany and stated that only 23 % of newly diagnosed patients with chronic lower back pain (CLBP) receive guideline-based multimodal therapy in the first year of the occurrence [21]. A big gap between evidence and practice in treating BP still exists [15, 22]. This gap significantly increases medical treatment costs without being associated with major treatment success among patients [23].

To address this gap, patient-centred, multimodal care programmes are proactively offered by health care providers directly to those affected by LBP [24, 25]. Although there are validated screening possibilities on-site to assess the need for treatment [26–30], a targeted and tailored approach during the first contact by the payer is more effective from the client's and payer's point of view. On the one hand, the individual concerned feels directly addressed and understands exactly the content of the offer, which leads to an increased motivation to participate [31, 32]. The payer, on the other hand, directly selects clients (on the basis of claims data), who are not expected to improve naturally and therefore focuses his efforts on potential beneficiaries, i.e., those likely receiving a medical benefit through an intervention. A medical

improvement will potentially result in an economic advantage. Targeted client selection would make it possible to implement and offer interventions that are dominant in terms of health economics - i.e., an improvement in health status on the one hand and a reduction in direct medical costs on the other [11, 33, 34].

Selecting the target group of chronic clients from purely administrative data is however not easily performed, as there is no classification by chronicity in the ICD-10 system. No distinction is made between acute, subacute and chronic pain. Most BP is simply classified as ‘*Low back pain*’ (an M54.5 diagnosis) [35].

To address this deficit, Freytag et. al. developed an algorithm based on routine data from an SHI in Germany with 5.2 million beneficiaries to classify BP patients before the invitation. They performed a secondary data analysis and used the data, which was originally intended for medical settlements, to build a classification tool [36].

As a result, they were able to divide patients into three chronicity classes (CC): 1: *without evidence of chronicity*, 2: *evidence of risk of chronicity*, 3: *evidence of chronicity*. However, they did not validate their assessment with actual patient feedback or patient-reported outcome, so that external validity is missing to a certain degree [37]. This issue led to the research question of the accuracy of the prediction model. Does the application classify patients correctly? Can it predict the actual chronicity according to von Korff’s GCPS from administrative data? And is a targeted client selection and thus more effective and efficient care possible with this tool?

For the investigation of this research question routine data from the private health insurance (PHI) provider Generali Deutschland Krankenversicherung AG (Generali Germany Health Insurance, formerly known as “Central Krankenversicherung”) was analysed. Since 2014, Generali has been running a proactively offered, multidisciplinary biopsychosocial rehabilitation (MBR) intervention for clients with CLBP [24, 38]. GCPS at enrolment, as well as the classification in CC according to Freytag et al., was available for participants.

The primary objective of this study was to assess the criterion validity of the predictive model. In doing so, medical and economic differences between the two models (GCPS vs. CC) were identified. This led to the secondary objective of an updated representation of the costs of care for CLBP patients in Germany. With that decision-makers could obtain a meaningful estimate of the annual expected costs of the treatment of CLBP patients in a PHI setting in Germany. Thus, they are able to prioritise actions and treatment decisions in a more informed way to determine how to best use limited resources.

Methods

Study Design

This was a cross-sectional study. Participants of the study were insured members of the Generali Deutschland Krankenversicherung AG, who signed up for long-term MBR against CLBP between July 2014 and March 2021.

All MBR participants underwent a digital assessment at the beginning of the intervention – the information to calculate the GCPS was collected here among others. Two ways to sign up for the programme existed. The standard way (I) was an invitation sent out by the insurance company based on the specific disease history as stated below. The alternative path (II) was based on the clients' initiative, where they directly requested participation in the health programme (further referred to as *self-selected*). For the invited insurance holder (I), the CC was available at the date of invitation. This was calculated on the basis of the submitted medical bills of the last 12 months before the invitation. The CC of insured persons who enrolled within three months of invitation was compared with the GCPS at the point of enrolment. For the secondary outcome of the cost analysis, participants who proactively requested participation (II) were additionally taken into account. Data management and statistical analyses were carried out using the software R[39] and the listed packages [40–43].

Participants

Invited participants (I) were selected according to CC [37]. For the calculation of the CC, the following routinely collected data from 12 months prior to invitation, were taken into account:

- Number of BP specific ICD-10 diagnoses (M40 – M54)
- Incapacity to work due to BP and its duration
- Use of strong opioids (ATC Group: N02A) as an indication of chronic pain and
- Psychiatric ICD-10 F-diagnoses: F32*, F33*, F34.1, F34.8, F34.9, F38, F41.2, F45.4, F48.0, F43.20, F43.21, F43.22, F54, F62.80

The three chronicity classes were assigned as:

1) Without evidence of chronicity:

Two M40 to M54 diagnoses and not CC group 2 or 3.

2) Evidence of risk of chronicity:

Two M40 to M54 diagnoses combined with less than two opioid prescriptions and either a) incapacity to work due to an M40 to M54 diagnosis of less than six weeks or b) at least two F diagnoses.

3) Evidence of chronicity:

Two M40 to M54 diagnoses combined with either a) incapacity to work of at least six weeks or b) at least two opioid prescriptions within six months.

As the insurance company subdivided participants by BP severity in the digital assessment and assigned a suitable programme variant, persons with all three CC levels were invited. Therefore, the minimum requirement to be invited was the presence of two ICD-10 diagnoses in the range of M40 to M54 within the last 12 months. Excluded from invitation were individuals with any condition that precluded participation in an intensive physical intervention (e.g., stroke or need for care). The complete exclusion list can be consulted elsewhere [24].

The aim of the study was to validate the CC algorithm. To achieve this, the GCPs was used [10] and compared with the CC. Participants were questioned about the duration, intensity and impairment due to their BP within the previous six months prior to the date of enrolment. Depending on the answers to the seven questions, every participant was assigned a GCPs grade. Grades ranged from:

- 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.

Variables

Primary Outcome

The primary outcome was the criterion validity of the CC, i.e. an evaluation of the accuracy of the prediction model of BP chronicity classes using claims data as developed by Freytag and colleagues. The GCPs was used as a reference value for the classification of chronicity. The predicted (CC) was compared with the actual chronicity grade (GCPs) for all invited participants of the MBR. To compare the four-level GCPs with the three-level CC, the GCPs needed to be reduced by one grade. GCPs grades I and II were combined and compared with *CC 1 - "without evidence of chronicity"*. GCPs grade III was compared with *CC 2 - "Evidence of risk of chronicity"* and GCPs grade IV with *CC 3 - "evidence of chronicity"*.

In a first step, the correlation between CC and newly categorised GCPs was assessed using Spearman's rho rank correlation coefficient with 95 % confidence intervals (CI). Strength of correlation was interpreted as *weak* ($\rho < 0.1$), *modest* ($\rho 0.1 - 0.3$), *moderate* ($\rho 0.31 - 0.5$), *strong* ($\rho 0.51 - 0.8$) or *very strong* ($\rho > 0.8$) [44]. The second step included the assessment of the agreement between CC and the categorised GCPs by using Cohen's weighted Kappa. The agreement was interpreted as *poor* (Kappa < 0.2), *fair* (Kappa $0.21 - 0.4$), *moderate* (Kappa $0.41 - 0.6$), *substantial* (Kappa $0.61 - 0.8$) or *almost perfect* (Kappa $0.81 - 1$) [45].

Furthermore, the GCPs and the CC were both dichotomised in severe and non-severe BP cases. Grades I and II were previously defined as *functional* chronic pain, and Grades III and IV as *non-functional* chronic pain [10]. In order to allow easier comparability and interpretation, GCPs grades I and II, which were already summarised, were relabelled as *non-severe* and III to IV as *severe* cases. CC class 1 and 2 equally as *non-severe*, and CC 3 as *severe* BP cases and presented in a 2x2 confusion matrix.

The confusion matrix assigned the chronicity class of each MBR participant with its predicted class (severe BP or non-severe BP). As a result, every sample belonged to one of the following four classes:

- *True positive (TP)* were actual severe BP cases that were correctly predicted as severe
- *True negative (TN)* were actual non-severe BP cases that were correctly predicted as non-severe
- *False positive (FP)* were actual non-severe BP cases that were wrongly predicted as severe
- *False negative (FN)* were actual severe BP cases that were wrongly predicted as non-severe

Sensitivity (i.e. the proportion of participants with severe BP who were correctly classified by the model), specificity (i.e. the proportion of participants without severe BP correctly classified as not having severe BP by the model) and Matthews correlation coefficient (MCC) [46] (i.e. the correlation between actual and predicted severity grades) were estimated to evaluate the model's performance. MCC was chosen instead of accuracy and F_1 score as it is more reliable taking into account all of the four confusion matrix categories [47]. As MCC is a discrete case of Pearson Correlation Coefficient, the strength of correlation was interpreted equally, meaning: *very weak relation* (MCC 0.01 – 0.29), *fair relation* (MCC 0.3 – 0.59), *moderately strong relation* (MCC 0.6 - 0.79) or *very strong relation* (MCC $\geq$ 0.8) [48]. Cohen's weighted Kappa was again estimated as a concordance statistic.

$$\text{Sensitivity} = \frac{\text{Severe BP cases correctly identified using claims data}}{\text{Total severe BP cases in MBR}} \times 100$$

$$\text{Specificity} = \frac{\text{Non severe BP cases correctly identified using claims data}}{\text{Total non severe BP cases in MBR}} \times 100$$

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}}$$

Participant characteristics potentially associated with the grade of BP chronicity

Demographic information of the participants (e.g. age, sex), overall health (e.g. weighted Charlson Comorbidity Index Score (CCI) [49], self-assessed overall health status using the first item of SF-12), possible psychological comorbidities (PHQ-4 score and its subscales [50], ICD-10 F-diagnoses) and direct effects of BP (ICD-10 M-diagnoses, everyday impairment, average pain level, number of days restricted in everyday activities within the last six months) were selected. These variables were descriptively compared across CC respective GCPS grades.

Not every participant was enrolled in a daily sickness benefit insurance in addition to their regular PHI policy at this provider. It is likely that most participants were insured against sick leave at another provider. However, no information was available on the insurance status. Therefore, in contrast to the SHI system, there was no general incentive for the insured to report incapacity to work to Generali. Since the days of incapacity to work played a major role in the calculation of the CC, the daily sickness allowance insurance status of the insured was regarded as a possible confounder and analysed separately in a sensitivity analysis. However, it was assumed that those insured against sick leave at this provider also reported absence.

Secondary outcome

The secondary outcome was an updated representation of the costs of care for CLBP in the German PHI setting. Overall health costs and BP specific inpatient, as well as outpatient costs in the last 12 months before enrolment, were considered. Costs were descriptively compared across CC respective GCPS grades.

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. Additional elective services (e.g., one or two-bedroom supplement) and the entire costs of dental care treatment were excluded.

In a PHI setting the reimbursement procedure follows the principle of refund of expenses, i.e., the clients pay the health care bill in advance, submits the bill afterwards to their insurance company and receives the reimbursement according to the insurance tariff concluded from it. Reimbursement of the health care bills depends on the respective tariff. The study population consisted of fully insured participants with different levels of deductible as well as policyholders eligible for governmental aid. Therefore, the cost component was defined as the total bill amount instead of the refund amount paid. Thus, the actual costs were compared with each other without taking into account which payer (health insurance, subsidy or individual supplementary) reimbursed the costs. As costs were presented for a period of 12 months, no discounting was executed. All costs were converted to 2020 Euros (€) using consumer price indices.

As healthcare costs tend to be highly skewed and heavily right-tailed [51], a truncated mean was also calculated in addition to the average costs per category. For this, all upper outliers (high-cost cases) were calculated using Tukey's method with $1,5 \times$ interquartile range (IQR) [52]. Low-cost cases were defined as participants who did not submit an invoice from the presented area in the last 12 months before enrolment.

Data Source/Measurement

For this study, two data sources were used. The information to calculate the CC and its connected variables (e.g., diagnoses, sick-days, opioid use, CCI), as well as all cost data, were obtained from claims data of the insurance. The information to calculate the GCPS was obtained through participants' responses in the standardised, self-administered digital assessment during enrolment. Participants were questioned about their current health status to a) assign the best type of intervention and b) control for individual developments with follow-up measurements.

Bias

The routinely collected data did not yield a potential source of bias. For the data collected within the standardised, self-administered questionnaire there were two potential sources of a) recall bias and b) demand characteristics. A recall bias was possible since the GCPS was calculated using the development of BP within the last six months. However, the GCPS is in general widely used to assess CLBP [8, 11, 53, 54] but also other types of anatomically defined pain conditions [55, 56]. It has been validated several times [57, 58] and is an internationally recognised tool in self-administered pain assessment [59] so that a possible effect of recall bias was neglected.

A second possible source of bias was demand characteristics [60], i.e. that respondents answer the questionnaire tactically in order to receive the most comprehensible care possible. However, participants were asked to answer truthfully in order to receive an intervention tailored to their individual needs. Since all participants were pain patients, who volunteered for the intervention, which is always free of charge, it could be assumed that their answers were rather accurate. Moreover, the specific steering logic was not mentioned in writing. Therefore, a bias due to demand characteristics seemed also unlikely.

Study Size

Different samples were required to answer the two research questions. The selection criteria are shown in Table 1. The study population consisted of 3,629 participants for whom the GCPS grade at enrolment was available. As the data was provided by a PHI, there were participants with an individual yearly threshold of costs before payment of expenses (deductible). Insurees with a fixed deductible usually only hand in their invoices of a year if they exceed that amount. To reduce the potential bias introduced by the

tariff, 122 participants who did not hand in any invoice in the 12 months before enrolment (yearly average of invoices = 27) were excluded.

To answer the first research question of the criterion validity all participants of the MBR who enrolled in the standard way (n = 2,722) were taken into account. The time between the initial invitation and enrolment was calculated. As CC was only available at the date of the invitation, participants who took longer than 90 days to register were excluded (n = 326) in order to rule out the temporal effect and thus potential changes of the CC. The final group size for the first research question was 2,396.

For the estimation of the cost of CLBP participants signed up on their own initiative were additionally considered (n = 872). The size of the group used to answer the second research question increased to 3,506.

Table 1 - Data preparation processes for selection of study population

Data processing steps showing the number of participants	Overall	Used in
<i>Study size</i>	3629	
<i>Exclusion of participants without any billing invoice available</i>	3506	Research question II
<i>Enrolment after invitation by insurance</i>	2722	
<i>Enrolment within 90 days after invitation</i>	2396	Research question I
<i>Enrolment within 90 days after invitation plus insured against sick leave</i>	1114	Sensitivity analysis

Results

Participants

Characteristics of the study population are presented in Table 2. Participants' mean age was 54.74 years, whereas 65.9 % were of male and 34.1 % of female sex. The mean CCI was 0.8, indicating a population in a healthy state. This was confirmed at the assessment. More than two-thirds self-reported an overall health status of "moderate" or better. The average PHQ-4 sum score was 2.82. The PHQ-4 subscales averaged 1.56 in the depression and 1.26 in the anxiety part. The average pain intensity was 4.48 and the average disability was 3.98, whereas both were assessed using the GCPS-questionnaire. Most participants reported being less than 14 days disabled due to their BP within the last six months. Nearly half (46.1 %) of the study population was insured against sick leave at this provider. In total, 8.7 % handed in a claim for sick leave due to BP.

A clear division of the characteristics by the GCPS could be observed. There was a clear negative correlation between GCPS and health status. The higher the grade, the lower the health indicators. This

was also true for variables, which were not used to calculate the GCPS (i.e., PHQ-4, overall health, CCI and sick leave due to BP).

The majority of participants were grouped in GPCS Grades I (42.8 %) or II (24.7 %). Grade III (17.7 %) and IV (14.8 %) made up one-third of the participants. Characteristics were also divided by CC in comparison (Appendix 1). It could be observed that CC 2 (evidence of risk of chronicity) reported on average a higher PHQ-4 than CC 1, CC 3 or self-selected participants.

Table 2 - Characteristics of study participants based on Graded Chronic Pain Grades

	Overall	GCPS I	GCPS II	GCPS III	GCPS IV	p
<i>N</i>	3506	1499	867	622	518	
		(42.8)	(24.7)	(17.7)	(14.8)	
<i>Sex = Female (%)</i>	1196 (34.1)	477 (31.8)	328 (37.8)	204 (32.8)	187 (36.1)	0.016
<i>Age (mean (SD))</i>	54.73 (9.52)	54.63 (9.56)	54.00 (9.59)	55.55 (9.44)	55.28 (9.30)	0.009
<i>CCI – score (mean (SD))</i>	0.80 (1.41)	0.71 (1.36)	0.72 (1.29)	0.88 (1.46)	1.06 (1.63)	<0.001
<i>Overall health (%)</i>						<0.001
<i>very good</i>	49 (1.4)	41 (2.7)	2 (0.2)	2 (0.3)	4 (0.8)	
<i>good</i>	505 (14.4)	340 (22.7)	105 (12.1)	41 (6.6)	19 (3.7)	
<i>moderate</i>	1832 (52.3)	916 (61.1)	490 (56.5)	279 (44.9)	147 (28.4)	
<i>bad</i>	970 (27.7)	200 (13.3)	250 (28.8)	261 (42.0)	259 (50.0)	
<i>very bad</i>	150 (4.3)	2 (0.1)	20 (2.3)	39 (6.3)	89 (17.2)	
<i>PHQ-4 score (mean (SD))</i>	2.82 (2.56)	1.76 (1.83)	2.75 (2.18)	3.79 (2.65)	4.83 (3.14)	<0.001
<i>PHQ-4 subscale depression</i>	1.56 (1.41)	0.94 (1.01)	1.53 (1.23)	2.12 (1.43)	2.70 (1.63)	<0.001
<i>PHQ-4 subscale anxiety</i>	1.26 (1.37)	0.82 (1.04)	1.22 (1.25)	1.67 (1.46)	2.12 (1.71)	<0.001
<i>Average pain intensity within the last six months (mean (SD))</i>	4.48 (1.95)	2.95 (1.32)	5.33 (1.26)	5.52 (1.63)	6.23 (1.62)	<0.001
<i>Average disability within the last six months (mean (SD))</i>	3.99 (2.44)	2.17 (1.58)	4.03 (1.64)	5.62 (1.70)	7.18 (1.47)	<0.001
<i>Days disabled (%)</i>						<0.001
<i>0 - 6 days</i>	2160 (61.6)	1353 (90.3)	710 (81.9)	97 (15.6)	0 (0.0)	
<i>07 - 14 days</i>	445 (12.7)	121 (8.1)	147 (17.0)	176 (28.3)	1 (0.2)	
<i>15 - 30 days</i>	398 (11.3)	22 (1.5)	10 (1.2)	268 (43.1)	98 (18.9)	

31 - 180 days	503 (14.3)	3 (0.2)	0 (0.0)	81 (13.0)	419 (80.9)
---------------	---------------	---------	---------	--------------	---------------

Definition of GCPS: Grade I: low disability-low intensity; Grade II: low disability-high intensity; Grade III: high disability-moderately limiting; Grade IV: high disability-severely limiting

Figure 1 depicts the distribution of GCPS grades per CC. More than every second participant (51.6 %) with a prediction of CC 1 belongs to GCPS grade I, 27.3 % to GCPS II, 13.4 % to GCPS III and 7.6 % to GCPS IV. The majority of participants predicted to belong to CC 2 are grouped to non-severe BP (GCPS grade I: 42.1 %, II: 19.4 %). More than half of the predicted CC 3 members belong to the severe BP group (GCPS grade III: 22.7 %, grade IV: 35.8 %). Self-selected participants are distributed across all categories (GCPS grade I: 32.8 %, II: 29.8 %, III: 20.5 %, IV: 16.9 %).

Criterion Validity

The primary outcome was to evaluate the criterion validity of the algorithm to define chronicity classes according to Freytag et al. [37]. The predicted CC was compared with the actual chronicity grade (GCPS) for all invited participants of the MBR. To compare the GCPS with the CC, the GCPS was reduced from four to three grades. In a first step, a 3x3 confusion matrix was created (see appendix). A total of 2,396 participants were categorised into one of the three classes (*chronic, risk of chronification or non-chronic BP*). With a value of 0.343 (95 % CI: 0.307 – 0.377, p<0.001) in Spearman’s rho, the correlation between the classes could be classified as *moderate*. A weighted Kappa of 0,307 (CI 0.268 – 0.346) indicated a *fair* agreement between CC and GCPS.

Table 3 - Evaluation of the predicted BP severity with self-reported GCPS grades

<i>Predicted BP severity category</i>	<i>Observed BP severity category</i>		
	<i>Severe</i>	<i>Non-severe (GCPS I & II)</i>	<i>Total</i>
<i>Severe (CC 3)</i>	TP: 401	FP: 392	793
<i>Non-severe (CC 1&2)</i>	FN: 333	TN: 1270	1603
<i>Total</i>	734	1662	2396
<i>Sensitivity</i>	54.6 %		
<i>Specificity</i>	76.4 %		
<i>Correctly predicted</i>	69.7 %		
<i>MCC</i>	0.304		
<i>Cohen's weighted Kappa</i>	0.304 (95 % CI: 0.260 – 0.348)		

GCPS = Graded Chronic Pain Grade [10] based on self-questionnaire at enrolment

CC = Chronicity Class [37] based on administrative claims data

TP = True positive were actual severe BP cases that were correctly predicted as severe

TN = True negative were actual non-severe BP cases that were correctly predicted as non-severe

FP = False positive were actual non-severe BP cases that were wrongly predicted as severe

FN = False negative were actual severe BP cases that were wrongly predicted as non-severe

Further, the GCPS and the CC were dichotomised in severe BP or non-severe BP cases. Table 3 shows the confusion matrix, which matches the assigned CC of each MBR participant with its predicted class. Results show a sensitivity of 54.6 % and a specificity of 76.4 %. In total 69.7 % were correctly predicted. With an MCC of 0.304, the strength of correlation was classified as *fair*. This was in agreement with Cohen's weighted Kappa of 0.304 (95 % CI: 0.260 – 0.348) – which also indicated a *fair* agreement between CC and GCPS.

Sensitivity Analysis

The capacity to work played a pivotal role in the calculation of CC. Only 46.1 % of the study population were insured against sick leave at this provider, which meant that only information about working ability was available for that subpopulation. To exclude the possibility of confounding due to the insurance status, a sensitivity analysis was run including only participants who were insured against sick leave (n = 1,114). In a similar fashion as shown above, a 3x3 confusion matrix (appendix) was created and afterwards dichotomised. Spearman's rho (0.405 (0.355 – 0.453, p<0.001)) and Cohen's weighted Kappa (0.358 (CI 0.298 – 0.418)) were increased. The 2x2 confusion matrix (Table 4) shows that sensitivity has

risen to 63.9 % whereas specificity has slightly reduced to 73.4 %. Overall, 70.7 % of all cases were correctly predicted. An MCC of 0.348 and a weighted Kappa of 0.341 also indicated a *fair* relationship between prediction and observation of BP severity for the subgroup.

Table 4 – Sensitivity analysis with participants insured against sick leave

<i>Predicted BP severity category</i>	<i>Observed BP severity category</i>		
	Severe	Non-severe (GCPS I & II)	Total
	(GCPS III & IV)		
<i>Severe (CC 3)</i>	TP: 202	FP: 212	414
<i>Non-severe (CC 1&2)</i>	FN: 114	TN: 586	700
<i>Total</i>	316	798	1114
<i>Sensitivity</i>	63.9 %		
<i>Specificity</i>	73.4 %		
<i>Correctly predicted</i>	70.7 %		
<i>MCC</i>	0.348		
<i>Cohen's weighted Kappa</i>	0.341 (95% CI: 0.275 – 0.408)		

GCPS = Graded Chronic Pain Grade [10] based on self-questionnaire at enrolment

CC = Chronicity Class [37] based on administrative claims data

TP = True positive were actual severe BP cases that were correctly predicted as severe

TN = True negative were actual non-severe BP cases that were correctly predicted as non-severe

FP = False positive were actual non-severe BP cases that were wrongly predicted as severe

FN = False negative were actual severe BP cases that were wrongly predicted as non-severe

Healthcare Costs of chronic BP patients

The secondary outcome was an updated representation of the costs of care for CLBP in the German PHI setting. Overall health costs and BP specific inpatient, as well as outpatient costs in the last 12 months before enrolment, were presented. Costs were descriptively compared across CC respective GCPS grades of participants who either were invited by the insurance or took part upon self-selection.

Table 5 presents the cost data of the 3,506 participants analysed. The pattern seen in the general characteristics of the study population could also be identified in the healthcare usage and cost information. A clear linear relationship between GCPS grade and presented variables could be observed. Healthcare usage and costs rose with increasing GCPS grades. The overall average of total direct health

costs was €7,279.78. The 1,499 participants with GCPS grade I had a mean cost of €5,967.94 whereas the 518 participants with GCPS grade IV at enrolment had a mean cost of €10,619.29. The average BP specific cost for the overall group was €1,082.13. Participants with GCPS grade IV (€2,312.12) had more than three times higher costs than GCPS grade I participants (€650.53). The truncated mean of BP specific costs was €751.73 for the overall group. The cost range between the mean of GCPS I (€618.80) and IV (€1,051.59) participants was considerably smaller. Skewness was reduced from 7.2 in the mean to 1.1 in the truncated mean. Cost differences could be mainly explained by high-cost cases. The distribution of high-cost cases varied significantly between groups: GCPS grade I group showed 3.9 % and group IV 24.9 % high-cost cases.

Similarly, the number of low-cost cases (i.e., participants who did not submit a BP invoice in the last 12 months before enrolment) showed a clear negative trend in combination with GCPS. In total 851 participants^[1] did not submit a BP-related invoice in the last 12 months. Nearly one-third of those belong to GCPS grade I (31.7 %) and 13.7 % belong to GCPS grade IV.

A grouping of the costs was also done by the CC (see Appendix). It could be observed that CC 3 is highly comparable with GCPS grade IV. The average total health costs for CC 3, which involved 519 participants, were €10,277.41. BP specific costs amount to €2,453.11 on average. Of applicable sick-leave participants, 69.5 % had to call in sick due to BP in the last 12 months. A high comorbidity with psychological disorders could be observed in CC 2. An ICD-10 F-diagnosis was available for 78 % of CC 2 – respectively 5.8 % for CC 1 and 15.4 % for CC 3. Participants whose participation was initiated upon self-selection (n = 803) could be grouped in between CC 2 and CC 3.

Table 5 - Healthcare usage and costs during the 12 months before enrolment based on GCPS

	Overall	GCPS I	GCPS II	GCPS III	GCPS IV	p	Skew
<i>N (%)</i>	3506	1499 (42.8)	867 (24.7)	622 (17.7)	518 (14.8)		
<i>Insured against sick-leave</i>	1616 (46.1)	744 (49.6)	396 (45.7)	260 (41.8)	216 (41.7)	0.001	
<i>Sick-leave due to BP (% of insured against sick-leave)</i>	305 (18.9)	67 (9)	35 (8.8)	72 (27.7)	131 (60.6)	<0.001	
<i>F-Diagnosis available (%)</i>	644 (18.4)	209 (13.9)	135 (15.6)	152 (24.4)	148 (28.6)	<0.001	
<i>Amount of ICD-10 F-Diagnoses (mean (SD))</i>	1.42 (4.84)	0.95 (3.63)	1.16 (4.55)	1.90 (5.81)	2.60 (6.56)	<0.001	5.9
<i>Amount of ICD-10 M-Diagnoses (mean (SD))</i>	4.11 (5.67)	2.70 (4.06)	3.45 (4.11)	5.16 (6.66)	8.04 (8.07)	<0.001	3.0
Total Health Cost € (mean (SD))	7279.78 (8040.56)	5967.94 (6579.83)	6570.73 (6646.93)	8648.47 (10151.72)	10619.29 (9787.68)	<0.001	3.6
<i>High-cost cases (%)</i>	252 (7.2)	75 (5.0)	48 (5.5)	57 (9.2)	72 (13.9)	<0.001	
BP Total Cost € (mean (SD))	1082.13 (2298.10)	650.53 (1531.07)	875.17 (1348.64)	1378.91 (2711.49)	2321.12 (3857.27)	<0.001	7.2
<i>BP truncated Cost (mean (SD))</i>	751.73 (670.23)	618.80 (617.22)	733.09 (634.29)	851.71 (682.29)	1051.59 (754.24)	<0.001	1.1
<i>High-cost cases (%)</i>	325 (9.3)	58 (3.9)	70 (8.1)	68 (10.9)	129 (24.9)	<0.001	
<i>Low-cost cases (%)</i>	851 (24.3)	475 (31.7)	194 (22.4)	111 (17.8)	71 (13.7)	<0.001	
BP Inpatient Cost – € (mean (SD))	340.70 (1822.82)	160.56 (1146.28)	157.26 (834.00)	463.77 (2114.65)	1021.22 (3398.49)	<0.001	10.5
<i>High-Cost cases (%)</i>	256 (7.3)	54 (3.6)	44 (5.1)	57 (9.2)	101 (19.5)	<0.001	
<i>Low-Cost</i>	3250	1446	823	565 (90.8)	417	<0.001	

<i>cases (%)</i>	(92.7)	(96.4)	(94.9)		(80.5)		
<i>BP Outpatient</i>	741.43	489.97 (818.02)	717.91 (927.41)	915.14 (1176.70)	1299.90 (1364.68)	<0.001	2.7
<i>Cost – € (mean (SD))</i>	(1047.45)						
<i>High-cost cases (%)</i>	216 (6.2)	41 (2.7)	52 (6.0)	45 (7.2)	78 (15.1)	<0.001	
<i>Low-cost cases (%)</i>	863 (24.6)	480 (32.0)	197 (22.7)	111 (17.8)	75 (14.5)	<0.001	

Definition of GCPS: Grade I: low disability-low intensity; Grade II: low disability-high intensity; Grade III: high disability-moderately limiting; Grade IV: high disability-severely limiting

High-Cost Cases were calculated using Tukey's method with $1,5 * IQR$ [52]

Truncated mean: Exclusion of high-cost cases and cases who did not submit a BP invoice in the last 12 months before enrolment.

[1] To be selected by CC two ICD-10 M-diagnoses within the last 12 months were the minimum requirement. In the presented table, the time difference between initial invitation by insurance and enrolment was not taken into consideration. 624 participants had two BP specific diagnoses (and connected invoices) within the last 12 months of invitation, but not within the last 12 months before enrolment. Additionally, 227 participants self-selected themselves and also did not hand in a BP specific invoice in the last 12 months before enrolment.

Discussion

Summary

This study used administrative claims data and self-reported patient information to validate a claims-based algorithm (CC) identifying the severity of CLBP. A functioning algorithm would enable payers to select and invite participants for targeted, expensive treatment programs without the need for additional screening. Results showed a *fair* correlation between predicted CC and actual GCPS grades. A total of 69.7 % of all cases was classified correctly. Sensitivity and specificity rates of 54.6 % and 76.4 % underlined the accuracy of the prediction. A sensitivity analysis with participants insured against sick leave showed similar results. The correlation between CC and GCPS with an MCC of 0.348 and a weighted Kappa of 0.341 indicated also a *fair* relationship between prediction and observation of BP severity for the subgroup.

Cost data could be clearly grouped by GCPS grades. It could be stated that the higher the grade, the higher the cost and health care usage. Overall, the average total direct health cost was €7,279.78. Participants with GCPS grade I had mean costs of €5,967.94 whereas participants with GCPS grade IV had mean costs of €10,619.29. The average BP specific cost for the overall group was €1,082.13.

Participants with GCPS grade IV (€2,312.12) had more than three times higher BP specific costs than GCPS Grade I participants (€650.53).

Limitations

The study had two limitations: the data used (I) and the outcome (II).

The administrative data used had the primary purpose of settling claims and was granted by a PHI, who in general have the freedom of implementing and offering health programmes without any restrictions due to national regulations. However, PHI data do not include all health-related billing data. Tariff-related peculiarities (e.g., deductibles and co-payments) mean that in practice not all medical invoices are submitted [61]. With an exclusion of participants with no invoices submitted in the last 12 months, possible tariff biases were reduced. But still, only a minority of participants held a daily sickness benefits insurance with this provider. An underreporting of sick leave was likely. The sensitivity analysis focused on participants who were insured against sick leave showed an improvement in the strength of correlation. However, it could be the case that the algorithm is better suited for a sickness fund where complete information about sick leave for all participants would be available (e.g., the German SHI). In further research, participants should be questioned about sick leave directly to cross-validate claims data.

The second limitation was the reduction of classes in the outcome of criterion validity. Original GCPS had four grades, whereas CC had only three grades. Therefore, GCPS grades I and II were combined in order to reduce the amount to three grades. Taking cost information and demographic characteristics into consideration it could be stated that this was a legitimate operation, as characteristics between I and II only differed slightly. A dichotomisation of the GCPS in severe and non-severe BP cases was also unproblematic as this is inherently contained in the grading, which is separating grades I/II and III/IV by disability into functional and non-functional chronic pain. A dichotomisation of the three CC classes could prove to be difficult, as CC 2 was between chronic and non-chronic. However, confusion matrices were run for 3x3 and 2x2 comparisons and outcomes differed only marginal. The strength of the relationship remained in the range of *a fair correlation* so that a reduction to two categories did not influence the overall results.

Interpretation

To our knowledge, this was the first study to compare the predicted BP severity by claims data with the actual BP severity by GCPS. In other types of diseases, predictive models from administrative data were often used to estimate disease severity. Studies predicting the severity of asthma [62], cancer [63], COPD [64] or stroke [65] identified comparable levels of performance. A correct classification of 70 % seems to be a reasonable accuracy to identify BP severity. Based on the findings of this study, the usage of the CC as a single tool to determine who gets treated against CLBP at what intensity is not recommended. A sensitivity of 54.6 % means that almost half of the participants who are experiencing

severe consequences based on GCPS would not be selected by the model. In addition, the insured persons' preferences and personal life circumstances should be considered to a reasonable extent when selecting the suitable program component. When creating further prediction algorithms, future research should include a criterion validity assessment in the validation study. Only then is it possible to actually determine whether the algorithm identified also performs well in practice.

It was also shown in this study that participants with high GCPS grades are not only suffering the most but also causing the highest costs. In previous studies it was clearly established that especially high GCPS grades profit from multimodal, long-term interventions [11, 24, 53]. The payer therefore has an interest in trying to focus on this subgroup to reach a cost-effective intervention.

Usage of the CC by Freytag could help to shape the type of intervention the individual receives. From cost data analysis it became clear that participants with CC 3 definitely should be targeted for intervention. Participants with CC 2 showed high psychological comorbidities with an average of 7.21 F-diagnoses in the last 12 months. The mean PHQ-4 of 3.78 (1.92 on the depression and 1.86 on the anxiety subscale) at enrolment was the highest compared to all other groups. Even though the subscale averages for the group were below the disorder cut point of ≥ 3 [50], they still indicated a need for special focus on psychological concerns for certain individuals exceeding the barrier. Several studies have focused on the connection between the spine and the brain and identified promising solutions like pain neuroscience education or biofeedback that could be applied especially to this target group of CC 2 [66–68]. At least participants with CC 1 should be additionally classified by a clinical screening item like STarT Back or GCPS before the appropriate intervention is assigned. As the predictive validity of both tools is highly comparable [69], decision-makers should select one of both based on item simplicity and ease of administration.

The second outcome of the study was an updated view of direct healthcare costs of patients with CLBP. For the German insurance market in which the study was conducted, two studies existed, which depicted healthcare costs of CLBP patients by chronicity grades [11, 12]. Wenig et. al [12] used a postal survey and asked SHI-insured participants with BP (n = 5,650) about their healthcare usage in the previous 3 months. From this, they estimated and extrapolated BP specific direct (46 %) and indirect costs (54 %) for 12 months. They estimated that a patient with CLBP would on average create direct costs of €612.50. They found out that the most influential predictor of high costs was a high GCPS grade. Participants with GCPS grades IV (€7,115.7) were said to have more than 17 times total costs (direct and indirect BP specific) than participants with grade I (€414.4).

In this study – which only focused on direct health costs – we also saw a sharp increase in BP specific costs based on GCPS grade. We used actual claims data from a 12 months period before the enrolment date in a health program against BP. Participants with grade IV (€2,321.12) had about 3.5 higher BP specific direct healthcare costs than participants with grade I (€650.53). With an average of €1,082.13 on BP specific total costs, the expenses in the PHI were higher than identified by Wenig et al. The difference between grades I and IV were albeit not as high as reported by Wenig and colleagues.

Müller et. al presented a study in 2019 in which they compared the therapeutic and economic effects of a multimodal back exercise programme. They also presented direct medical costs for a study population of 2,324 participants using routine data supplied by an SHI. Participants with GPCS grade IV (€5,310) had 2.2 times higher overall direct healthcare costs than participants with grade I (€2,391) over a time period of two years.

In the presented study we identified 1.8 times higher overall direct health costs (€10,619 vs. €5,968) in the last 12 months. Total direct costs of the privately insured were albeit a lot higher than costs in the SHI system. This can be explained by the fact that reimbursement schemes and provider spending in the outpatient setting tend to be two to three times higher in the PHI setting [70, 71].

Nevertheless, this presented study gives a good overview and updated cost information on direct overall and BP specific costs based on their different GPCS grades. As the relationship of the cost differences between different GPCS grades was in accordance with previous studies, it can be assumed that the figures presented are a good representation of the costs to be expected in the PHI system. Decision-makers should use these findings to match effective interventions with limited funds. Participants with GPCS grade IV are suffering greatly and produce the highest costs through the current treatment of their pain. As care is often not guideline-based [14, 15, 22, 23], a great deal of attention should be paid to targeting this cohort. Involving them in appropriate and effective treatment programmes is essential.

Generalisability

The presented study had three strengths that played a role in the generalisability of the results: I: study sample, II: target group and III: availability of the data. Due to the long time period of seven years in which data was collected, the study reached a large size with 3,506 participants. Furthermore, only data of participants was used who felt their BP so pressing that they were willing to participate in a long-term intervention. The monetary key figures shown thus reflect expected costs of people participating in an intervention against their BP.

One other advantage was the availability of the data. In order to participate, it was mandatory to carry out the digital assessment, so that a lot of information about BP and its consequences could be obtained. In addition, the cost data routinely collected from the insurance company could be used purposefully.

Due to the setting in a PHI, the generalisability of the presented cost data is nonetheless limited. All participants were insured with the same PHI company. Even though recruitment took place nationwide, outpatient cost data and therefore also the overall cost was probably still higher than could be expected in a comparable study focused on the SHI. This was due to the systemic differences between SHI and PHI and cannot be remedied. With the knowledge of the two to three times higher costs in the outpatient sector, trends could albeit also be gained for the SHI. The trend in the spending between different GPCS grades was however highly comparable between PHI and SHI. Moreover, inpatient spending was highly comparable between both systems, as additional private elective benefits, such as supplements for

treatment by a chief physician or accommodation in a single room, were excluded from the cost consideration. Besides, the inpatient reimbursement scheme of diagnosis-related groups (DRG) is identical in both systems.

It could also be possible that in an SHI setting, with complete data about coverage against sick leave, more than 70 % of the participants would be correctly categorised by the CC. However, the sensitivity analyses with the subgroup of sick leave insured showed that the predictive ability only improved slightly. The strength of correlation still stayed in the range of a *fair* agreement between CC and GCPS so that the influence of the system here could be regarded as low. Overall, the strengths of the study outweigh the inherent disadvantages of PHI claims data so that the results can be interpreted and transferred to other settings. If the prediction algorithm of Freytag et al. is to be used in other settings, care should be taken to ensure that information on diagnoses and medications as well as on work absences due to BP and its duration is available and reliable.

Conclusion

This was the first study to compare predicted BP severity by claims data with actual BP severity by GCPS. They result in a sensitivity of 54.6 % and a specificity of 76.4 %. In total 69.7 % were correctly predicted. With an MCC of 0.304, the strength of correlation was classified as *fair*. Based on the findings of this study, the usage of CC as a single tool to determine who gets treated against CLBP and with what is not recommended. Healthcare spending clearly can be separated by GCPS. A predictive algorithm that could abolish the need for medical screening on site would need to reach a very high sensitivity to identify patients who would profit the most from a targeted intervention (GCPS grades III and IV). The CC by Freytag is a good tool to segment candidates for BP specific interventions – especially to identify possible participants who need additional psychological components in the intervention. However, it cannot replace a self-reported medical screening instrument. The rate of false negatives would be too high for that.

Abbreviations

Abbreviation	Explanation
BP	Back Pain
CC	Chronicity Class
CCI	Charlson's Comorbidity Index Score
CI	Confidence Intervals
CLBP	Chronic lower back pain
DRG	Diagnosis-related groups
FN	False negative were actual severe BP cases that were wrongly predicted as non-severe
FP	False positive were actual non-severe BP cases that were wrongly predicted as severe
GCPS	Graded Chronic Pain Scale
GP	General practitioner
ICD-10	10th revision of the International Statistical Classification of Diseases and Related Health Problems
IQR	Interquartile range
LBP	Low back pain
MBR	Multidisciplinary biopsychosocial rehabilitation
MCC	Matthews correlation coefficient
PHI	Private Health Insurance
PHQ-4	Patient Health Questionnaire 4
SHI	Statutory Health Insurance
TN	True negative were actual non-severe BP cases that were correctly predicted as non-severe
TP	True positive were actual severe BP cases that were correctly predicted as severe

Declarations

Ethics approval and consent to participate: The study used routinely collected data from an intervention, which was evaluated and registered previously at the German Clinical Trials Register under DRKS00015463 retrospectively (4 Sept 2018). Consent to participate and the self-reported questionnaire have remained unchanged since the study and are therefore still valid in accordance with the ethics proposal.

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 already consented to the usage of the data for further analysis, no new ethic vote was sought for the present analysis.

Written informed consent was obtained from all study participants.

Consent for publication: not applicable, no individual data.

Availability of data, material and code: The data and code that support the findings of this study are available from Generali Deutschland Krankenversicherung AG but restrictions apply to the availability of these data, which were used under license for the current study and so are not publicly available. They are however available from the authors upon reasonable request and with permission including a signed data access agreement of Generali Deutschland Krankenversicherung AG.

Competing interests: Martin Hochheim (MH) is working in part time for the Generali Health Solutions GmbH (GHS), which is affiliated with the Generali Deutschland Krankenversicherung AG. 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.

Funding: In this study, data from the medical digital enrolment questionnaire for a health programme for insured with BP were linked with administrative data to carry out the analysis. The medical programme was funded by the Generali Deutschland Krankenversicherung AG. Funding for this study by any party has not taken place.

Authors' contributions: MH planned the analyses, analysed the data, interpreted the results and wrote the manuscript. VA supervised the project. MW contributed to the implementation of the research. 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.

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

References

1. Wu, A., March, L., Zheng, X., Huang, J., Wang, X., Zhao, J., Blyth, F.M., Smith, E., Buchbinder, R., Hoy, D.: Global low back pain prevalence and years lived with disability from 1990 to 2017: estimates from the Global Burden of Disease Study 2017. *Ann. Transl. Med.* 8, 299–299 (2020). <https://doi.org/10.21037/atm.2020.02.175>
2. Hartvigsen, J., Hancock, M.J., Kongsted, A., et al.: What low back pain is and why we need to pay attention. *Lancet.* 391, 2356–2367 (2018). <https://doi.org/https://doi.org/10.1016/S0140->

3. van Tulder, M., Becker, A., Bekkering, T., Breen, A., Gil del Real, M.T., Hutchinson, A., Koes, B., Laerum, E., Malmivaara, A.: Chapter 3 European guidelines for the management of acute nonspecific low back pain in primary care. *Eur. Spine J.* 15, s169–s191 (2006). <https://doi.org/10.1007/s00586-006-1071-2>
4. Urits, I., Burshtein, A., Sharma, M., Testa, L., Gold, P.A., Orhurhu, V., Viswanath, O., Jones, M.R., Sidransky, M.A., Spektor, B., Kaye, A.D.: Low Back Pain, a Comprehensive Review: Pathophysiology, Diagnosis, and Treatment, <https://doi.org/10.1007/s11916-019-0757-1>, (2019)
5. Maher, C., Underwood, M., Buchbinder, R.: Non-specific low back pain. *Lancet.* 389, 736–747 (2017). [https://doi.org/10.1016/S0140-6736\(16\)30970-9](https://doi.org/10.1016/S0140-6736(16)30970-9)
6. James, S.L., Abate, D., Abate, K.H., 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.* 392, 1789–1858 (2018). [https://doi.org/10.1016/S0140-6736\(18\)32279-7](https://doi.org/10.1016/S0140-6736(18)32279-7)
7. Von Korff, M., Miglioretti, D.L., Korff, M. Von, Miglioretti, D.L.: A prognostic approach to defining chronic pain. *Pain.* 117, 304–313 (2005). <https://doi.org/10.1016/j.pain.2005.06.017>
8. Turner, J.A., Shortreed, S.M., Saunders, K.W., Leresche, L., Berlin, J.A., Von Korff, M.: Optimizing prediction of back pain outcomes. *PAIN®.* 154, 1391–1401 (2013). <https://doi.org/10.1016/J.PAIN.2013.04.029>
9. Ritzwoller, D.P., Crounse, L., Shetterly, S., Rublee, D.: The association of comorbidities, utilization and costs for patients identified with low back pain. *BMC Musculoskelet. Disord.* 7, 72 (2006). <https://doi.org/10.1186/1471-2474-7-72>
10. Von Korff, M., Ormel, J., Keefe, F.J., Dworkin, S.F.: Grading the severity of chronic pain. *Pain.* 50, 133–149 (1992). [https://doi.org/10.1016/0304-3959\(92\)90154-4](https://doi.org/10.1016/0304-3959(92)90154-4)
11. Müller, G., Pfander, M., Clement, M., et al.: Therapeutic and economic effects of multimodal back exercise: A controlled multicentre study. *J. Rehabil. Med.* 51, 61–70 (2019). <https://doi.org/10.2340/16501977-2497>
12. Wenig, C.M., Schmidt, C.O., Kohlmann, T., Schweikert, B.: Costs of back pain in Germany. *Eur. J. Pain.* 13, 280–286 (2009). <https://doi.org/10.1016/j.ejpain.2008.04.005>
13. Gatchel, R.J., Polatin, P.B., Noe, C., Gardea, M., Pulliam, C., Thompson, J.: Treatment- and Cost-Effectiveness of Early Intervention for Acute Low-Back Pain Patients: A One-Year Prospective Study. *J. Occup. Rehabil.* 13, 1–9 (2003). <https://doi.org/10.1023/A:1021823505774>
14. Foster, N.E., Anema, J.R., Cherkin, D., et al.: Prevention and treatment of low back pain: evidence, challenges, and promising directions. *Lancet.* 391, 2368–2383 (2018). [https://doi.org/10.1016/S0140-6736\(18\)30489-6](https://doi.org/10.1016/S0140-6736(18)30489-6)
15. Corp, N., Mansell, G., Stynes, S., Wynne-Jones, G., Morsø, L., Hill, J.C., van der Windt, D.A.: Evidence-based treatment recommendations for neck and low back pain across Europe: A systematic review of guidelines. *Eur. J. Pain (United Kingdom).* 25, 275–295 (2021). <https://doi.org/10.1002/ejp.1679>

16. Searle, A., Spink, M., Ho, A., Chuter, V.: Exercise interventions for the treatment of chronic low back pain: a systematic review and meta-analysis of randomised controlled trials. *Clin. Rehabil.* 29, 1155–1167 (2015). <https://doi.org/10.1177/0269215515570379>
17. Steffens, D., Maher, C.G., Pereira, L.S.M., Stevens, M.L., Oliveira, V.C., Chapple, M., Teixeira-Salmela, L.F., Hancock, M.J.: Prevention of Low Back Pain. *JAMA Intern. Med.* 176, 199 (2016). <https://doi.org/10.1001/jamainternmed.2015.7431>
18. van Middelkoop, M., Rubinstein, S.M., Kuijpers, T., Verhagen, A.P., Ostelo, R., Koes, B.W., van Tulder, M.W.: A systematic review on the effectiveness of physical and rehabilitation interventions for chronic non-specific low back pain. *Eur. Spine J.* 20, 19–39 (2011). <https://doi.org/10.1007/s00586-010-1518-3>
19. 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.* 8, 871 (2019). <https://doi.org/10.3390/jcm8060871>
20. Niemier, K.: Multimodal, polypragmatisch und kostenintensiv. *Man. Medizin.* 50, 16–27 (2012). <https://doi.org/10.1007/s00337-011-0888-x>
21. Daniel, T., Koetsenruijter, J., Wensing, M., Wronski, P.: Chronische Kreuzschmerzen – Nutzertypen ambulanter Versorgung. *Der Schmerz.* (2021). <https://doi.org/10.1007/s00482-021-00565-2>
22. Hall, A.M., Scurrey, S.R., Pike, A.E., Albury, C., Richmond, H.L., Matthews, J., Toomey, E., Hayden, J.A., Etchegary, H.: Physician-reported barriers to using evidence-based recommendations for low back pain in clinical practice: A systematic review and synthesis of qualitative studies using the Theoretical Domains Framework. *Implement. Sci.* 14, 49 (2019). <https://doi.org/10.1186/s13012-019-0884-4>
23. Werber, A., Schiltenswolf, M.: Treatment of Lower Back Pain—The Gap between Guideline-Based Treatment and Medical Care Reality. *Healthcare.* 4, 44 (2016). <https://doi.org/10.3390/healthcare4030044>
24. 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.* 19, 714 (2019). <https://doi.org/10.1186/s12913-019-4561-8>
25. Müller-Schwefe, G., Freytag, A., Höer, A., Schiffhorst, G., Becker, A., Casser, H.R., Glaeske, G., Thoma, R., Treede, R.D.: Healthcare utilization of back pain patients: Results of a claims data analysis. *J. Med. Econ.* 14, 816–823 (2011). <https://doi.org/10.3111/13696998.2011.625067>
26. Langenmaier, A.M., Amelung, V.E., Karst, M., Krauth, C., Püschner, F., Urbanski, D., Schiessl, C., Thoma, R., Klasen, B.: Subgroups in chronic low back pain patients – A step toward cluster-based, tailored treatment in inpatient standard care: On the need for precise targeting of treatment for chronic low back pain. *GMS Ger. Med. Sci.* 17, (2019). <https://doi.org/10.3205/000275>
27. Jonsdottir, S., Ahmed, H., Tómasson, K., Carter, B., Tomasson, K., Carter, B., Tómasson, K., Carter, B.: Factors associated with chronic and acute back pain in Wales, a cross-sectional study. *BMC Musculoskelet. Disord.* 20, 215 (2019). <https://doi.org/10.1186/s12891-019-2477-4>

28. Hill, J.C., Dunn, K.M., Lewis, M., Mullis, R., Main, C.J., Foster, N.E., Hay, E.M.: A primary care back pain screening tool: Identifying patient subgroups for initial treatment. *Arthritis Rheum.* 59, 632–641 (2008). <https://doi.org/10.1002/art.23563>
29. Karstens, S., Krug, K., Hill, J.C., Stock, C., Steinhäuser, J., Szecsenyi, J., Joos, S.: Validation of the German version of the STarT-Back Tool (STarT-G): A cohort study with patients from primary care practices Rehabilitation, physical therapy and occupational health. *BMC Musculoskelet. Disord.* 16, (2015). <https://doi.org/10.1186/s12891-015-0806-9>
30. Karran, E.L., McAuley, J.H., Traeger, A.C., Hillier, S.L., Grabherr, L., Russek, L.N., Moseley, G.L.: Can screening instruments accurately determine poor outcome risk in adults with recent onset low back pain? A systematic review and meta-analysis. *BMC Med.* 15, 13 (2017). <https://doi.org/10.1186/s12916-016-0774-4>
31. Forberger, S., Bammann, K., Bauer, J., et al.: How to tackle key challenges in the promotion of physical activity among older adults (65+): The AEQUIPA network approach. *Int. J. Environ. Res. Public Health.* 14, (2017). <https://doi.org/10.3390/ijerph14040379>
32. Zubala, A., MacGillivray, S., Frost, H., Kroll, T., Skelton, D.A., Gavine, A., Gray, N.M., Toma, M., Morris, J.: Promotion of physical activity interventions for community dwelling older adults: A systematic review of reviews. *PLoS One.* 12, (2017). <https://doi.org/10.1371/journal.pone.0180902>
33. Almazrou, S.H., Elliott, R.A., Knaggs, R.D., Alaujan, S.S.: 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>
34. Herman, P.M., Lavelle, T.A., Sorbero, M.E., Hurwitz, E.L., Coulter, I.D.: 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>
35. Juniper, M., Le, T.K., Mladsi, D.: The epidemiology, economic burden, and pharmacological treatment of chronic low back pain in France, Germany, Italy, Spain and the UK: a literature-based review. *Expert Opin. Pharmacother.* 10, 2581–2592 (2009). <https://doi.org/10.1517/14656560903304063>
36. Swart, E., Gothe, H., Geyer, S., Jaunzeme, J., Maier, B., Grobe, T.G., Ihle, P.: Gute Praxis Sekundärdatenanalyse (GPS): Leitlinien und Empfehlungen. *Gesundheitswesen.* 77, 120–126 (2015). <https://doi.org/10.1055/s-0034-1396815>
37. Freytag, A., Thiede, M., Schiffhorst, G., Höer, A., Wobbe, S., Luley, C., Glaeske, G.: Cost of Back Pain and the Significance of Chronic Pain - Results of a Claims Data Analysis. *Gesundheitsökonomie und Qual.* 17, 79–87 (2012). <https://doi.org/10.1055/s-0031-1281578>
38. 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.* 81, 831–838 (2017). <https://doi.org/10.1055/s-0043-121696>

39. 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.
40. Wickham, H., Averick, M., Bryan, J., et al.: Welcome to the Tidyverse. *J. Open Source Softw.* 4, 1686 (2019). <https://doi.org/10.21105/joss.01686>
41. Grolemund, G., Wickham, H.: Dates and Times Made Easy with lubridate. *J. Stat. Softw.* 40, 1–25 (2011). <https://doi.org/10.18637/jss.v040.i03>
42. 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
43. Signorell, A., Aho, K., Alfons, A., et al.: DescTools: Tools for descriptive statistics. R package version 0.99.41. (2021). <https://cran.r-project.org/package=DescTools>. Accessed May 19, 2021
44. Muijs, D.: Doing Quantitative Research in Education with SPSS. SAGE Publications Ltd, 1 Oliver’s Yard, 55 City Road, London EC1Y 1SP United Kingdom (2011)
45. Landis, J.R., Koch, G.G.: The Measurement of Observer Agreement for Categorical Data. *Biometrics.* 33, 159 (1977). <https://doi.org/10.2307/2529310>
46. Matthews, B.W.: Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *BBA - Protein Struct.* 405, 442–451 (1975). [https://doi.org/10.1016/0005-2795\(75\)90109-9](https://doi.org/10.1016/0005-2795(75)90109-9)
47. Chicco, D., Jurman, G.: The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics.* 21, 6 (2020). <https://doi.org/10.1186/s12864-019-6413-7>
48. Powers, D.M.W.: Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *Int. J. Mach. Learn. Technol.* 2, 37–63 (2011). <https://doi.org/10.9735/2229-3981>
49. Charlson, M.E., Pompei, P., Ales, K.L., MacKenzie, C.R.: A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. *J. Chronic Dis.* 40, 373–383 (1987). [https://doi.org/10.1016/0021-9681\(87\)90171-8](https://doi.org/10.1016/0021-9681(87)90171-8)
50. Kroenke, K., Spitzer, R.L., Williams, J.B.W., Löwe, B.: An ultra-brief screening scale for anxiety and depression: The PHQ-4. *Psychosomatics.* 50, 613–621 (2009). <https://doi.org/10.1176/appi.psy.50.6.613>
51. Mihaylova, B., Briggs, A., O’Hagan, A., Thompson, S.G.: Review of statistical methods for analysing healthcare resources and costs. *Health Econ.* 20, 897–916 (2011). <https://doi.org/10.1002/hec.1653>
52. Tukey, J.W.: Exploratory Data Analysis. Addison-Wesley Publishing Company (1977)
53. Müller, G., Lyssenko, L., Giurgiu, M., Pfander, M., Clement, M., Kaiserauer, A., Heinzl-Guntenbrunner, M., Kohlmann, T., Bös, K.: How effective and efficient are different exercise patterns in reducing back pain? *Eur. J. Phys. Rehabil. Med.* 56, 585–593 (2020). <https://doi.org/10.23736/S1973-9087.20.05975-4>

54. Ferrer-Peña, R., Calvo-Lobo, C., Aiguadé, R., Fernández-Carnero, J.: Which Seems to Be Worst Pain Severity and Quality of Life between Patients with Lateral Hip Pain and Low Back Pain. *Pain Res. Manag.* 2018, (2018). <https://doi.org/10.1155/2018/9156247>
55. Canales, G.D.L.T., Guarda-Nardini, L., Rizzatti-Barbosa, C.M., Conti, P.C.R., Manfredini, D.: Distribution of depression, somatization and pain-related impairment in patients with chronic temporomandibular disorders. *J. Appl. Oral Sci.* 27, (2019). <https://doi.org/10.1590/1678-7757-2018-0210>
56. Chantaracherd, P., John, M.T., Hodges, J.S., Schiffman, E.L.: Temporomandibular joint disorders' impact on pain, function, and disability. *J. Dent. Res.* 94, 79S-86S (2015). <https://doi.org/10.1177/0022034514565793>
57. Smith, B.H., Penny, K.I., Purves, A.M., Munro, C., Wilson, B., Grimshaw, W.J., Chambers, W.A., Smith, W.C.: The chronic pain grade questionnaire: Validation and reliability in postal research. *Pain.* 71, 141–147 (1997). [https://doi.org/10.1016/S0304-3959\(97\)03347-2](https://doi.org/10.1016/S0304-3959(97)03347-2)
58. Elliott, A.M., Smith, B.H., Smith, W.C., Chambers, W.A.: Changes in chronic pain severity over time: The chronic pain grade as a valid measure. *Pain.* 88, 303–308 (2000). [https://doi.org/10.1016/S0304-3959\(00\)00337-7](https://doi.org/10.1016/S0304-3959(00)00337-7)
59. Von Korff, M., DeBar, L.L., Krebs, E.E., Kerns, R.D., Deyo, R.A., Keefe, F.J.: Graded chronic pain scale revised: mild, bothersome, and high-impact chronic pain. *Pain.* 161, 651–661 (2020). <https://doi.org/10.1097/j.pain.0000000000001758>
60. Allen, M.: *The SAGE Encyclopedia of Communication Research Methods*. SAGE Publications, Inc, 2455 Teller Road, Thousand Oaks California 91320 (2017)
61. Gothe, H., Ihle, P., Matusiewicz, D., Swart, E.: *Routinedaten im Gesundheitswesen. Handbuch Sekundärdatenanalyse: Grundlagen, Methoden und Perspektiven*. Verlag Hans Huber, Bern (2014)
62. Birnbaum, H.G., Ivanova, J.I., Yu, A.P., Hsieh, M., Seal, B., Emani, S., Rosiello, R., Colice, G.L.: Asthma Severity Categorization Using a Claims-Based Algorithm or Pulmonary Function Testing. *J. Asthma.* 46, 67–72 (2009). <https://doi.org/10.1080/02770900802503099>
63. Smith, G.L., Shih, Y.C.T., Giordano, S.H., Smith, B.D., Buchholz, T.A.: A method to predict breast cancer stage using Medicare claims. *Epidemiol. Perspect. Innov.* 7, (2010). <https://doi.org/10.1186/1742-5573-7-1>
64. Macaulay, D., Sun, S.X., Sorg, R.A., Yan, S.Y., De, G., Wu, E.Q., Simonelli, P.F.: Development and validation of a claims-based prediction model for COPD severity. *Respir. Med.* 107, 1568–1577 (2013). <https://doi.org/10.1016/j.rmed.2013.05.012>
65. Sung, S.F., Hsieh, C.Y., Lin, H.J., Chen, Y.W., Chen, C.H., Kao Yang, Y.H., Hu, Y.H.: Validity of a stroke severity index for administrative claims data research: a retrospective cohort study. *BMC Health Serv. Res.* 16, 1–9 (2016). <https://doi.org/10.1186/s12913-016-1769-8>
66. Nijs, J., Clark, J., Malfliet, A., Ickmans, K., Voogt, L., Don, S., den Bandt, H., Goubert, D., Kregel, J., Coppieters, I., Dankaerts, W.: In the spine or in the brain? Recent advances in pain neuroscience applied in the intervention for low back pain. *Clin. Exp. Rheumatol.* 35 Suppl 1, 108–115 (2017)

67. Sielski, R., Rief, W., Glombiewski, J.A.: Efficacy of Biofeedback in Chronic back Pain: a Meta-Analysis. *Int. J. Behav. Med.* 24, 25–41 (2017). <https://doi.org/10.1007/s12529-016-9572-9>
68. Hoffman, B.M., Papas, R.K., Chatkoff, D.K., Kerns, R.D.: Meta-analysis of psychological interventions for chronic low back pain. *Heal. Psychol.* 26, 1–9 (2007). <https://doi.org/10.1037/0278-6133.26.1.1>
69. Von Korff, M., Shortreed, S.M., Saunders, K.W., Leresche, L., Berlin, J.A., Stang, P., Turner, J.A.: Comparison of back pain prognostic risk stratification item sets. *J. Pain.* 15, 81–89 (2014). <https://doi.org/10.1016/j.jpain.2013.09.013>
70. European Union: Germany: Country Health Profile 2017. *Eur. J. Public Health.* 28, (2017). <https://doi.org/10.1787/9789264283398-en>
71. Gerlinger, T., Burkhardt, W.: Vergütung privatärztlicher Leistungen, <https://www.bpb.de/politik/innenpolitik/gesundheitspolitik/72637/verguetung-privataerztlicher-leistungen>. Accessed May 2, 2021

Figures

Figure 1

Comparison of actual BP severity with prediction Comparison of GCPS at baseline with prediction from claims data algorithm according to Freytag et. al

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Appendix.docx](#)