

RESEARCH ARTICLE

Validation of the German version of the STarT-MSK-Tool: A cohort study with patients from physiotherapy clinics

Sven Karstens^{1*}, Jochen Zebisch², Johannes Wey³, Roger Hilfiker⁴, Jonathan C. Hill⁵

1 Department of Computer Science, Therapeutic Sciences, Trier University of Applied Sciences, Trier, Germany, **2** Physio Meets Science, Leimen, Germany, **3** Department of Computer Science, Formerly Therapeutic Sciences, Trier University of Applied Sciences, Trier, Germany, **4** School of Health Sciences, HES-SO Valais-Wallis, Leukerbad, Switzerland, **5** School of Medicine, Keele University, Staffordshire, United Kingdom

* karstens@hochschule-trier.de



Abstract

Background

The STarT-MSK-Tool is an adaptation of the well established STarT-Back-Tool, used to risk-stratify patients with a wider range of musculoskeletal presentations.

Objective

To formally translate and cross-culturally adapt the Keele STarT-MSK risk stratification tool into German (STarT-MSK_G) and to establish its reliability and validity.

Methods

A formal, multi-step, forward and backward translation approach was used. To assess validity patients aged ≥ 18 years, with acute, subacute or chronic musculoskeletal presentations in the lumbar spine, hip, knee, shoulder, or neck were included. The prospective cohort was used with initial data collected electronically at the point-of-consultation. Retest and 6-month follow-up questionnaires were sent by email. Test-retest reliability, construct validity, discriminative ability, predictive ability and floor or ceiling effects were analysed using intra-class correlation coefficient, and comparisons with a reference standard (Orebro-Musculoskeletal-Pain-Questionnaire: OMPQ) using correlations, ROC-curves and regression models.

Results

The participants' ($n = 287$) mean age was 47 (SD = 15.8) years, 51% were female, with 48.8% at low, 43.6% at medium, and 7.7% at high risk. With ICC = 0.75 (95% CI 0.69; 0.81) test-retest-reliability was good. Construct validity was good with correlations for the STarT-MSK_G-Tool against the OMPQ-Tool of $r_s = 0.74$ (95% CI 0.68, 0.79). The ability of the tool [comparison OMPQ] to predict 6-month pain and disability was acceptable with AUC = 0.77 (95% CI 0.71, 0.83) [OMPQ = 0.74] and 0.76 (95% CI 0.69, 0.82) [OMPQ = 0.72]

OPEN ACCESS

Citation: Karstens S, Zebisch J, Wey J, Hilfiker R, Hill JC (2022) Validation of the German version of the STarT-MSK-Tool: A cohort study with patients from physiotherapy clinics. PLoS ONE 17(7): e0269694. <https://doi.org/10.1371/journal.pone.0269694>

Editor: Fatih Özden, Mugla Sitki Kocman Universitesi, TURKEY

Received: September 10, 2021

Accepted: May 25, 2022

Published: July 1, 2022

Copyright: © 2022 Karstens et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Requests on data sharing and data access can be sent to the local data protection management of Trier University of Applied Sciences (datenschutz@hochschule-trier.de).

Funding: The author(s) received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

respectively. However, the explained variance (linear/logistic regression) for predicting 6-month pain (21% [OMPQ = 17%]/logistic = 29%) and disability (linear = 20%:[OMPQ = 19%]/logistic = 26%), whilst being comparable to the existing OMPQ reference standard, fell short of the a priori target of $\geq 30\%$.

Conclusions

The German version of the STarT-MSK-Tool is a valid instrument for use across multiple musculoskeletal conditions and is available for use in clinical practice. Comparison with the OMPQ suggests it is a good alternative.

Introduction

Musculoskeletal (MSK) disorders comprising pain in the region of the lower back, the neck, or osteoarthritis affecting the joints of the upper or lower extremities are among the leading causes of disability. These complaints often have a chronic course and their burden on individuals and society is large [1–3]. Due to aging populations, it is estimated that the prevalence of MSK conditions will further rise [4]. Typically, the majority of patients with these conditions are managed by general practitioners and physiotherapists [5]. Patient-reported measurement instruments are used by these clinicians to a varying degree, but there is a need for generic prognostic tools and risk stratification methods that are usable across a variety of body sites to facilitate targeted treatment decision-making [6, 7]. A back-specific instrument, specifically designed to establish the prognosis of patients in primary care is the Keele STarT-Back-Tool (Subgrouping for Targeted Treatment). It allocates patients to one of three prognostic subgroups (low, medium and high risk) in which they receive a risk-matched treatment [8–10]. This procedure has shown effectiveness and has been implemented in routine care in the UK [9, 11]. Internationally, the successful reproduction of the research results remains limited, although practitioners describe positive experiences in clinical practice [12, 13]. One criticism is that this tool is limited to patients with low back pain and a tool applicable to a broader group of musculoskeletal patients would have much greater appeal and be easier to implement [14].

Through a programme of research the STarT-Back-Tool has therefore been adapted and validated to produce the Keele STarT-MSK risk stratification tool (STarT-MSK) for use in a broader musculoskeletal patient population [15]. This approach has been supported by a recent umbrella review indicating that there are a number of common prognostic factors among patients with MSK-complaints including: worse baseline function, higher symptom/pain severity, worse mental well-being, more comorbidities, older age and higher body mass index [16]. Several translations of the STarT-MSK-Tool are available, but a German version did not exist and knowledge about its measurement properties is limited [17–21].

A translated version of the original STarT-MSK could support German physiotherapists, physicians or other health professionals to be able to provide risk-based stratified care for musculoskeletal disorders [14, 18]. Risk-based stratified care may help clinicians to better target treatments according to a patients' individual risk status, thereby maximising the benefits of care and reducing unnecessary treatments and costs [22, 23]. Consensus on the primary care management options relevant for each risk-group has been identified for UK primary care [24], but this may need to be adapted to the German context. Sowden et al. described various matched treatments for back pain ranging from one-off advice sessions for low risk patients to

more comprehensive solutions addressing patients with complex biopsychosocial prognostic factors for high risk patients [25]. To develop comparable procedures, recommendations for matched treatments were gathered for patients with MSK-conditions [24, 26, 27], and together with the tool were integrated by an international research group developing a web-app informing first contact clinicians in their clinical decision making [28].

The objective of this study was, to formally translate and cross-culturally adapt the STarT-MSK-Tool into German (German version: STarT-MSK_G). Moreover, we aimed to investigate its test-retest reliability, construct validity discriminative ability, predictive ability and floor or ceiling effects.

Methods

Design

A cohort study including a retest (t1) and a half-year follow-up (t2) in addition to the initial assessment (t0) was conducted. Patients were recruited from physiotherapy clinics (n = 7). The inclusion criteria were patients 18 years or older with acute, subacute or chronic complaints in the region of the lumbar spine, the hip or knee, the shoulder or the neck. The exclusion criteria were those with a known or suspected tumor, an acute inflammatory condition, recent musculoskeletal-related surgery (last six months) or trauma (last 3 months). German language skills had to be sufficient to complete the study questionnaires and participants had to provide written consent and their email and telephone details for follow-up purposes.

Initial data (t0) was collected electronically in the clinics via SoSci Survey [29]. The invitations to answer the t1- and t2-questionnaires were sent by email. To counter memory effects and at the same time minimize changes due to the natural course, a period of one week between t0 and t1 was aimed for [30]. To reduce drop-outs, patients who did not respond to a t1- or t2- invitation received a reminder after one week and were phoned after two weeks.

Ethical approval was granted by the Ethics Committee of Trier University of Applied Sciences, Computer Science/Therapeutic Sciences (registration ID: 01–2019). All patients gave their written informed consent for participation before enrollment in the clinics.

Translation and cross-cultural adaptation of the STarT-MSK-tool

The instrument validated in this study is the STarT-MSK_G. There are two versions of the STarT-MSK-Tool: (1st) a self-report version which was used in this study and (2nd) a clinical interview version. A copy of the instrument can be requested here: www.keele.ac.uk/startmsk.

The translation and cross-cultural adaptation was done according to internationally accepted guidelines and with permission for translation from the developers of the original version [31]. The translation committee consisted of three people (SK, JW, JCH). Of those two had extensive experience in cross-cultural adaptation [32, 33]. A coordinator collected and synthesized translations. Forward translations were carried out by three people with German mother tongue; one lay person and two physiotherapists. Two of these translators were German, and one of the physiotherapists was from Switzerland to facilitate cross-national validity. The three forward translations were synthesized into a final forward version by the coordinator. This version was sent back to the translators and comments were invited. The backward translations were done by two non-medical translators who were native speakers of English. The two backward translations were sent for discussion to the developers of the original English version. Very good conformity of the backward translations with the original version was shown, but item ten was revised changing 'pains' ('Schmerzen') to 'pain condition' ('Schmerzproblematik').

To check for acceptability and comprehension a pre-test was carried out with 10 patients from a German physiotherapy clinic. A Think-Aloud method was utilized, while the tool was completed [34]. Moreover, patients were asked open questions to determine if they experienced any problems with the tool. Due to grammatical reasons the German version of item eight begins with the time frame ('the last two weeks') and 'feeling down/depressed' follows. This was preferred by the participants of the pretests, after two alternatives were presented. A report describing the translation process and including the different translations was sent to the developers and the German version was confirmed. A copy of the German version can be requested here: www.keele.ac.uk/startmsk.

Reference instruments

To test for construct validity several reference instruments were added. Based on a formative model, both, the STarT-MSK like the OMPQ assess the risk for future pain and disability, using a set of items of known biopsychosocial risk factors [15, 35, 36]. Moreover, depending on the patients' complaints one of the following instruments were used to determine disability: German version of the Neck Disability Index (NDI) [37], Shoulder Pain and Disability Index (subscale disability, SPADI_{DIS}) [38], Roland Morris Disability Questionnaire (RMDQ) [39] or Western Ontario and McMaster Universities Osteoarthritis Index (subscale disability, WOMAC_{DIS}) [40]. Pain intensity was measured using the mean of three eleven-point box-scales for least, average (over the previous two weeks), and current pain [41, 42].

The STarT-MSK comprises of 10 items. The first item is an 11-point numeric pain rating scale. The other nine items have a dichotomous response option: yes/no. To calculate a sum-score, the items are recoded (item 1: 0–4 = 0 points, 5–6 = 1 point, 7–8 = 2 points, 9–10 = 3 points; item 2 to 9: yes = 1 point, no = 0 points). The final score is calculated by summarizing the point for all 10 items, with a possible total score ranging from 0 to 12. Based on cut-off points established for the original version, a total score of ≤ 4 points indicates low risk, a total score between 5–8 points medium risk, and ≥ 9 points high risk for persisting pain disability [15].

To determine the OMPQ score the sum of the five subscale means was computed resulting in a possible range from 0 to 50 points [35]. The RMDQ-score equals the number of the items checked positive by the patients and can range from 0 to 24 points [39]. To the NDI-score, each question adds 0 to 5 points to a total maximal sum-score of 50 which is transformed to percentages ranging from 0 to 100 [37]. The WOMAC_{DIS}-score was calculated by summarizing the item values, then divided by the number of items resulting in a range from 0 to 10 points [40]. The SPADI_{DIS}-score was calculated by summarizing the item values, then divided by the number of valid items, with maximally one non-valid item accepted. This also resulted in a score ranging from 0 to 10 points [38].

Statistical analyses

Descriptive statistics were calculated to characterize the study population and each subgroup. The baseline characteristics of the study participants are provided to allow interpretability of the study sample. Moreover, numbers on recruitment rate, drop-outs and missing data were described.

To investigate the **test-retest reliability** the intraclass correlation coefficient (ICC based on a two-way random effect, absolute agreement model (2.1)) was used. *An ICC above 0.50 was considered acceptable* [43]. Additionally, Cohen's Kappa for agreement on item level was calculated to further explain test-retest-reliability.

For convergent **construct validity** the STarT-MSK_G was related to the OMPQ. Spearman correlations were calculated for the time point t0. A priori a positive correlation was expected, with higher scores meaning worse prognosis on both instruments. The magnitude of the reported correlation coefficient was evaluated with a correlation of 0.1–0.3 considered to be small, >0.3–0.5 to be moderate, and greater than 0.5 to be large [44]. *At least a moderate correlation of greater than 0.4 was considered sufficient.* Additionally, to visually represent the correlation of the instruments, box and whisker plot graphs were presented using the OMPQ-score for each subgroup defined by the STarT-MSK_G score. Next to the relation with the OMPQ, coefficients (Spearman) for the correlation with the reference instruments for disability were calculated across the pain sites (NDI, RMDQ, WOMAC_{Dis}, SPADI_{Dis}). In comparison to the OMPQ lower correlations were expected.

Floor and ceiling effects were considered present if more than 15% of the responders achieved the lowest or highest possible score [45]. *It was expected that ≤ 15% of the responders would achieve the lowest or highest possible score.*

To assess STarT-MSK_G's **discriminative ability** ROC (receiver operating characteristic) curves with areas under the curves (AUC) and 95% confidence interval (CI) were computed [46]. The curves were calculated for poor physical status at t0 (RMDQ [39], NDI [37], WOMAC_{Dis} [40], SPADI_{Dis} [38]). Moreover, ROC curves with AUC were computed for pain intensity and disability for all patients. To determine if a patient was a 'case' on reference standard instruments, the individual's scores were compared to cut-off values defined in the literature: RMDQ ≥ 4 [47], NDI ≥ 15 [48], WOMAC_{Dis} ≥ 2.1 [49], SPADI_{Dis} ≥ 4.1 [50].

Adjectives that can be used to describe AUC-values have been proposed by Hosmer and Lemeshow with an AUC = 0.5 suggesting 'no discrimination', 0.7 to < 0.8 considered 'acceptable discrimination', 0.8 to 0.9 considered 'excellent discrimination' and >0.9 considered 'outstanding discrimination' [51]. *At least acceptable discrimination for disability was expected.*

To analyse the **predictive ability** the t0 score of the STarT-MSK_G was used as the predictor variable in univariate linear regression. *It was aimed to explain a proportion of at least 30% of variance in the outcome (disability and pain).* For comparison purposes the variance explained by the OMPQ was also calculated. Additionally, logistic regression analyses were performed. For dichotomization of disability the thresholds given above were used (see discriminative ability), for pain intensity the median was used (with 2.7 at t0 and 4.3 at t2 this fitted well with thresholds described in the literature [52, 53]). The R²-statistics (adjusted/ Nagelkerke) explaining the variance were evaluated. To test the calibration of the logistic prediction models, Spiegelhalter's z test was used [54, 55].

Next to regression analyses and in parallel to the procedure described for discriminative ability, areas under the curves (AUC) with 95% confidence intervals (CI) were calculated for STarT-MSK_G predicting dichotomised t2-outcomes (dichotomisation see discriminative ability). Moreover, to enable comparison, AUCs were calculated for OMPQ predicting dichotomised t2-outcomes.

Terwee et al. suggested a sample size of 50 patients for construct validity and reliability [45]. Therefore, to enable analyses for subgroups defined by diagnosis, while allowing a drop-out of 10% and considering an uneven distribution (estimated smallest subgroup with 20%), it was aimed to recruit 300 patients in total.

As significance level alpha = 5% was set. Analyses were performed using SPSS version 27.0 and R language and environment for statistical computing, version 4.0.0 [56].

Results

Consent for participation was given by 287 patients. The mean age was 47 (SD 15.8) years, and 51% were female, with overall 48.8% at low, 43.6% at medium and 7.7% at high risk. (Table 1). Non-consenters (n = 36) on average were 8.6 years (CI 95% 3.7, 13.5), they were older and more often female (66%). During the previous twelve weeks before t0, 64 patients (22.3%) reported having taken some sick leave. The t1 questionnaire was returned by 261 patients (91%), the t2 questionnaire by 246 patients (86%). Forty-five patients (16%) answered the questionnaires before the first contact with the therapist. The median number of contacts before answering the questions of the other 242 patients was 3 (IQR = 3).

Table 1. Characteristics of the study population.

	Total			Lower back			Neck			Shoulder			Hip or Knee		
	Mean	SD	n	Mean	SD	n	Mean	SD	n	Mean	SD	n	Mean	SD	n
Age	47.0	15.8	287	46.4	16.0	122	42.5	14.9	65	51.0	13.3	40	50.6	16.6	60
Average pain	4.5	1.8	287	4.6	1.8	122	4.4	1.7	65	4.4	1.9	40	4.4	1.9	60
STarT-MSK_G score	4.7	2.6	287	5.2	2.5	122	4.5	2.7	65	3.6	2.3	40	4.7	2.6	60
OMPQ score	16.4	8.0	286	17.4	8.4	121	16.4	8.1	65	12.2	6.0	40	17.0	7.2	60

		Total		Lower back		Neck		Shoulder		Hip or Knee	
		n	%	n	%	n	%	n	%	n	%
Gender	male	140	48.8%	60	49.2%	28	43.1%	22	55.0%	30	50.0%
	female	147	51.2%	62	50.8%	37	56.9%	18	45.0%	30	50.0%
Education	Professional qualification	43	15.2%	18	15.0%	8	12.5%	7	17.5%	10	16.9%
	Other	12	4.2%	4	3.3%	3	4.7%	2	5.0%	3	5.1%
	General education or lower	62	21.9%	31	25.8%	11	17.2%	6	15.0%	14	23.7%
	Higher education entrance qualification or higher	166	58.7%	67	55.8%	42	65.6%	25	62.5%	32	54.2%
Occupation	20h per week or more	162	57.2%	70	58.3%	40	62.5%	23	57.5%	29	49.2%
	Up to 20h per week	51	18.0%	22	18.3%	9	14.1%	9	22.5%	11	18.6%
	None, unemployed	7	2.5%	3	2.5%	3	4.7%	0	0.0%	1	1.7%
	None, retired	43	15.2%	15	12.5%	6	9.4%	7	17.5%	15	25.4%
	None, househusband or -wife	8	2.8%	3	2.5%	3	4.7%	0	0.0%	2	3.4%
	None of the previous	12	4.2%	7	5.8%	3	4.7%	1	2.5%	1	1.7%
General health	Excellent	6	2.1%	4	3.3%	1	1.5%	1	2.5%	0	0.0%
	Very good	50	17.5%	15	12.4%	10	15.4%	9	22.5%	16	26.7%
	Good	141	49.3%	58	47.9%	41	63.1%	20	50.0%	22	36.7%
	Fair	79	27.6%	37	30.6%	11	16.9%	10	25.0%	21	35.0%
	Poor	10	3.5%	7	5.8%	2	3.1%	0	0.0%	1	1.7%
Sickness certificate	No	223	77.7%	92	75.4%	42	64.6%	36	90.0%	53	88.3%
	Yes	64	22.3%	30	24.6%	23	35.4%	4	10.0%	7	11.7%
Duration of complaints	Up to 11 weeks	98	34.1%	41	33.6%	24	36.9%	10	25.0%	23	38.3%
	12 or more weeks	189	65.9%	81	66.4%	41	63.1%	30	75.0%	37	61.7%
MSK comorbidity	No	170	59.2%	78	63.9%	28	43.1%	23	57.5%	41	68.3%
	Yes	117	40.8%	44	36.1%	37	56.9%	17	42.5%	19	31.7%
STarT-MSK Risk-Group	Low	140	48.8%	50	41.0%	34	52.3%	27	67.5%	29	48.3%
	Medium	125	43.6%	60	49.2%	25	38.5%	12	30.0%	28	46.7%
	High	22	7.7%	12	9.8%	6	9.2%	1	2.5%	3	5.0%

OMPQ: Örebro Musculoskeletal Pain Questionnaire, STarT-MSK_G: STarT-MSK-Tool, German version.

<https://doi.org/10.1371/journal.pone.0269694.t001>

There were 122 patients with lower back complaints, 65 with neck, 40 with shoulder and 60 with hip/knee complaints. Thirty-six patients (12.5%) previously received surgery in the region of their complaints. Additional details on the characteristics of the study population are given in [Table 1](#).

The median time interval between t0 and t1 was 7 (IQR = 8) days and between t0 and t2 181 (IQR = 11) days. The follow-up questionnaires sent to the patients at t1 and t2 were returned by 261 (91%) and 246 (86%) of the participants, respectively. Non-responders at t1 on average were 6.5 years younger than responders, with a large confidence interval (CI 95% -0.7, 13.8) and were less often female (responders 53% female, non-responders 38%). Non-responders at t2 on average were 6.6 years younger than responders, with a confidence interval not including zero (CI 95% 0.9, 12.3) and were less often female (responders 52% female, non-responders 46%).

Test-retest-reliability

The ICC (t0 to t1) for the STarT-MSK_G was 0.75 (95% CI 0.69; 0.81) and therefore, is 'good'. For individual items the median κ was 0.58 (range 0.42 (item 9) to 0.72 (item 7 and 10)) ([Table 2](#)).

Construct validity

Correlations for the STarT-MSK_G-Tool against the OMPQ-Tool was $r_s = 0.74$ (95% CI 0.68, 0.79; convergent construct validity). A visual presentation of the correlation is given in [Fig 1](#). Correlation for the STarT-MSK_G-Tool against the disability measures consistently was lower, ranging from $r_s = 0.44$ to $r_s = 0.71$ (details displayed in [Table 3](#)).

Floor/ceiling effects

With 3.8% of patients having a STarT-MSK score of 0 points and 0.3% with the maximal score of twelve points, no floor or ceiling effects were observed.

Discriminative ability

The AUC for STarT-MSK_G ability to discriminate disability cases at initial contact was 0.77 (95% CI 0.72, 0.83), indicating 'acceptable' discrimination. The AUC for pain was 0.83 (95% CI

Table 2. Kappa coefficients of single item test-retest of the STarT-MSK_G.

	Kappa	95% CI	
		lower	upper
Item 1*	0.58	0.49	0.66
Item 2	0.53	0.42	0.63
Item 3	0.46	0.37	0.57
Item 4	0.63	0.51	0.72
Item 5	0.58	0.48	0.65
Item 6	0.53	0.44	0.63
Item 7	0.72	0.64	0.81
Item 8	0.64	0.54	0.75
Item 9	0.42	0.27	0.57
Item 10	0.72	0.61	0.78

CI: confidence interval

*: squared weights.

<https://doi.org/10.1371/journal.pone.0269694.t002>

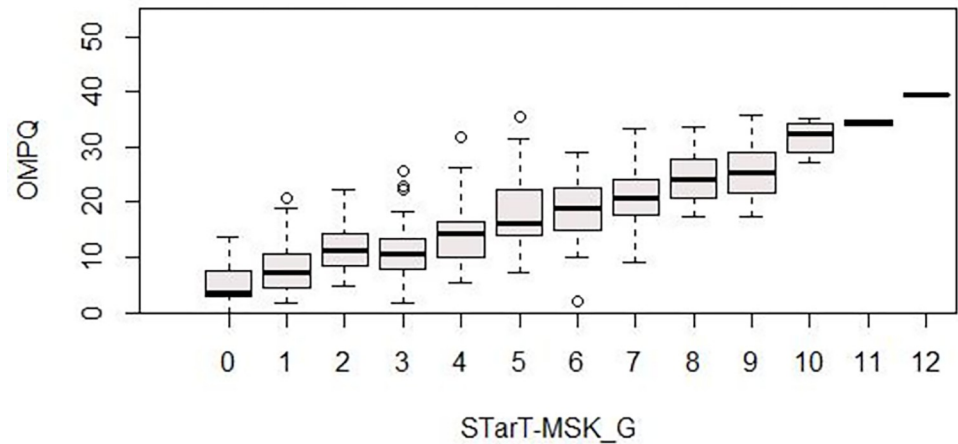


Fig 1. Relation between STarT-MSK_G and OMPQ. OMPQ: Örebro Musculoskeletal Pain Questionnaire, STarT-MSK_G: STarT-MSK-Tool, German version.

<https://doi.org/10.1371/journal.pone.0269694.g001>

0.78, 0.89), indicating ‘good’ discrimination (Fig 2). The AUCs for the different subgroups ranged from 0.68 to 0.85 (Table 4 and S1 Fig).

Predictive ability

Regression analyses. The univariate **linear** regression models statistical significance was seen with both $p < 0.001$, resulting in an amount of explained variance by the STarT-MSK_G of 21% (variance explained by OMPQ 17%) for pain at 6 months and 20% (variance explained by OMPQ 19%) for combined disability as shown by adjusted R^2 . The explained variance therefore fell short of the 30% aimed for, both for predicting pain and disability.

For the univariate **logistic** regression models, comparable were statistically significant, with both $p < 0.001$, resulting in an amount of variance explained by the STarT-MSK_G of 29% for pain at 6 months and 26% for combined disability as shown by Nagelkerke’s R^2 . With $z = -0.01$ for disability and $z = 0.07$ for pain, Spiegelhalter’s z was non-significant ($p = 0.99$; $p = 0.95$).

Areas under the curves

The AUC for STarT-MSK_G ability to predict pain-cases at follow-up was 0.77 (95% CI 0.71, 0.83). The AUC for disability was 0.76 (95% CI 0.70, 0.82), indicating ‘acceptable’ prediction (Fig 3). The AUC for OMPQ predicting pain-cases at follow-up was 0.74 (95% CI 0.68, 0.80)

Table 3. Correlations STarT-MSK_G against disability measures.

Reference	r_s	95% CI	
		Lower	Upper
RMDQ	0.589**	0.46	0.69
NDI	0.710**	0.57	0.81
SPADI_DIS	0.443**	0.12	0.67
WOMAC_DIS	0.536**	0.31	0.74

r_s : Spearman correlation coefficient; CI: confidence interval; RMDQ: Roland Morris Disability Questionnaire; NDI: Neck Disability Index; SPADI_DIS: Shoulder Pain and Disability Index, subscale disability; WOMAC_DIS: Western Ontario and McMaster Universities Osteoarthritis Index, subscale disability.

** significant correlation.

<https://doi.org/10.1371/journal.pone.0269694.t003>

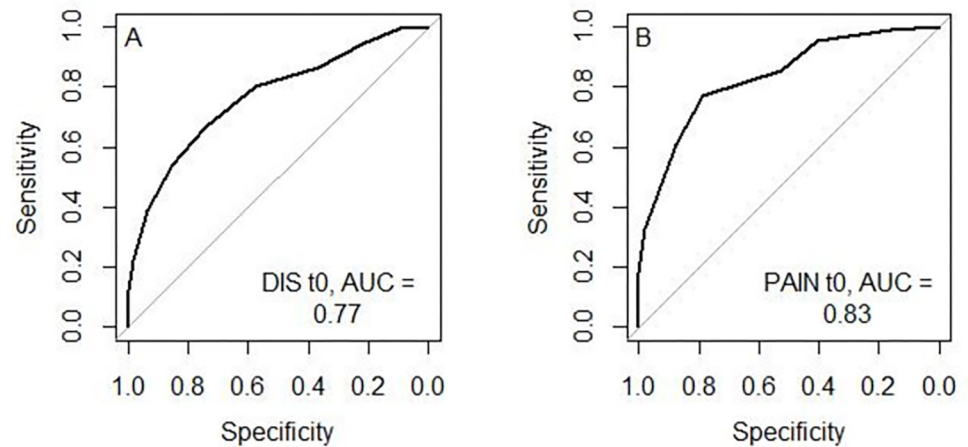


Fig 2. Receiver operating characteristic curves disability and pain at inclusion (t0). Combined disability score (DIS) and pain versus STarT-MSK_G.

<https://doi.org/10.1371/journal.pone.0269694.g002>

and for disability-cases 0.72 (95% CI 0.65, 0.78) (Fig 3). The AUCs for disability by subgroup ranged from 0.70 to 0.88 (Table 5 and S2 Fig), indicating overall ‘acceptable’ discrimination and ‘good’ discrimination for patients with hip or knee complaints.

Discussion

After cross-cultural adaptation of the STarT-MSK, a German version is now available and first information on its psychometric properties was established. Overall, these are promising, especially with good test-retest reliability and good construct validity. The instrument explained an amount of variance six months after the first measurement with slightly stronger predictive values than those for the OMPQ. Nevertheless, it fell short of the $\geq 30\%$ target.

To test the construct validity, the Örebro Musculoskeletal Pain Questionnaire (OMPQ) was used [35]. The good correlation between the instruments confirm that the STarT-MSK_G assesses risk for persisting pain disability. In comparison, correlations with instruments measuring disability was lower.

Predictive properties checked by ROC-Analyses resulted in acceptable AUCs that were higher than those of the OMPQ, although with extensively overlapping CI. The calculated, logistic and linear models with pain or disability outcomes explained from 20% to 29% of the variance in outcome, but did not exceed the pre-specified target of $R^2 > 30\%$. The number of 30% was estimated based on results from the development study, being unpublished at the time [15]. The achieved amount of explained variance fits very well to that from the external

Table 4. Areas under the curve (AUC) by subgroup at initial assessment.

Subgroup	AUC	95% CI	
		Lower	Upper
Lower back	0.77	0.68	0.85
Neck	0.85	0.76	0.94
Shoulder	0.68	0.50	0.86
Hip or knee	0.76	0.63	0.88

CI: confidence interval; AUC: area under the curve.

<https://doi.org/10.1371/journal.pone.0269694.t004>

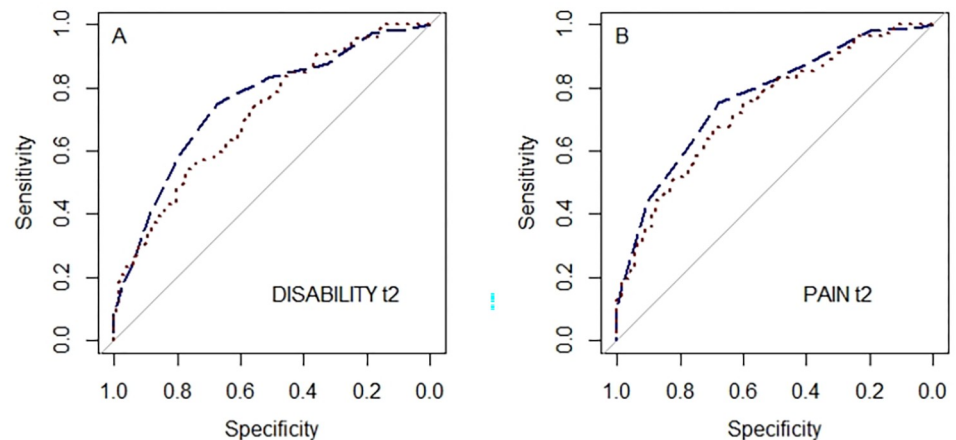


Fig 3. Receiver operating characteristic curves disability and pain at follow-up (t2). Combined DISABILITY score and PAIN versus STarT-MSK_G (dashed) and Örebro Musculoskeletal Pain Questionnaire (dotted).

<https://doi.org/10.1371/journal.pone.0269694.g003>

validation for the original version of the STarT-MSK [15]. In future studies the added value of the STarT-MSK_G together with covariables could be analysed for example in multiple regression analyses. Comparably, for the German version of the STarT-Back-Tool, adding a one item-variable capturing global health status and the baseline score of the outcome (disability) successfully increased the variance explained to $\geq 45\%$ [10]. Moreover, the suggestions given by Beneciuk et al. to use change-scores of the STarT-Tool might be considered [57]. In the present study the latter was not done, since the aim of data collection at t1 was to test retest-reliability. Van den Broek et al. just recently examined the predictive validity of the Dutch version of the STarT-MSK, by choosing a different statistical method [17]. Calculating relative risks, they showed that patients at low risk had a better prognosis than those at medium and especially than those at high risk. Major differences to that work—next to the language—are a shorter follow-up and a much smaller sample size leading to a total of three high-risk patients.

The treatment was not influenced by the researchers in the presented study, thus, it can be assumed that the content will have influenced the outcome at follow-up and the variance explained by the regression analyses. An alternative would have been to standardize the procedures, but such a shift from an observational to an experimental design would have led to costs exceeding the available resources for this project. On the other hand, to withhold therapy would have been unethical. Considering the other positive properties established for reliability and validity of the STarT-MSK_G, it would be worthwhile to develop a study design specifically aiming at improving prediction.

Table 5. Areas under the curve (AUC) by subgroup at follow-up.

Subgroup	AUC	95% CI	
		Lower	Upper
Lower back	0.70	0.60	0.79
Neck	0.76	0.63	0.89
Shoulder	0.75	0.58	0.94
Hip or knee	0.88	0.78	0.98

CI: confidence interval; AUC: area under the curve.

<https://doi.org/10.1371/journal.pone.0269694.t005>

All predictive ROC at least resulted in acceptable AUC. The cut-offs used to differ between cases and non-cases were derived from the literature [47–50]. Nonetheless, various methods exist to define cut-offs leading to different values [58]. A choice of different cut-offs might have resulted in diverging AUC.

The development of the STarT-MSK_G is related to that of the STarT-Back-Tool. While the area of application of STarT-Back-Tool is limited to low back pain [8], a strength of the STarT-MSK_G is its appropriability for patients with a variety of musculoskeletal complaints. In practice, such a possibility for generic use makes clinical processes easier with one instrument fitting for a broader group of patients. The administrative burden can be reduced, since often patients present with complaints at several sites simultaneously [59, 60]. Moreover, in future it might enable comparison of different patient-subgroups [61, 62].

Another strength of the STarT-MSK is that matched treatments were compiled and instruments assisting clinicians in decision-making are under development [24, 28, 63]. Such instruments help the clinician to address the patients' needs more specifically, eg. choosing cognitive behavioural-based approaches for patients at high risk of an unfavorable outcome [64, 65]. The knowledge of the measurement properties of the STarT-MSK_G strengthens its use for this purpose. Since practitioners have mixed ideas about how to best make use of prognostic tools [66, 67], strategies on how to best implement them should be further developed and detailed studies to describe the added value should be conducted [6].

Strength and weaknesses

Reference for the development of the STarT-MSK at Keele (UK) was the STarT-Back-Tool [8, 18]. Three of the authors (SK, RH, JCH) were involved in translation of the latter to German and testing of its psychometric properties, resulting in a valid version [32, 68]. The knowledge derived from this process, was an advantage for the work on STarT-MSK, since the researchers were familiar with the underlying concept.

Next to the German version multiple other translations were done including the Dutch, French, Hebrew and Norwegian versions which were validated. However, three of those studies were conducted with smaller sample-sizes and only one with a design enabling determination of the instrument's predictive ability [17, 19, 21]. For the Dutch version the predictive ability was confirmed, although the cited authors suggested a further external validation study [17].

The low amount of dropout in this study is an area of strength. It is substantially lower than the benchmark set by the Cochrane Back and Neck Group for long term follow-up [69]. This indicates that the developed strategy for data collection worked well, which is in line with descriptions on the low burden of online data collection [70].

Concerning the sample size, in the literature minimal numbers of 50 to 100 participants are required for validation [45, 71]. In total this was easily met and the lower number was also met for the different pain sites except for patients with shoulder complaints. Results for this subgroup should be considered preliminary and confirmed in future works, probably in a specifically tailored study, since these patients are most difficult to recruit [60]. The number of included high-risk patients, who are often seldom, even was comparably high, especially when considering the physiotherapeutic setting [17, 72].

Conclusion

The German version of the STarT-MSK-Tool is a valid instrument for use across multiple musculoskeletal conditions and is available for use in clinical practice. It fulfils the fundamental requirements for an assessment instrument having shown good test-retest-reliability, face,

construct validity and predictive validity when analysing ROC-Curves. The instrument explains a considerable amount of variance in six month pain and disability scores. However, whilst the prognostic abilities are comparable to those of the existing reference standard (OMPQ), as the variance was lower than the target set a priori, it is recommended that future research should seek to raise the predictive abilities of this tool further.

Supporting information

S1 Fig. Receiver operating characteristic curves disability by subgroup. Disability-scores versus STarT-MSK_G total score; RMDQ: Roland Morris Disability Questionnaire; NDI: Neck Disability Index; SPADI: Shoulder Pain and Disability Index, subscale disability; WOMAC: Western Ontario and McMaster Universities Osteoarthritis Index, subscale disability; t0: initial.

(TIF)

S2 Fig. Receiver operating characteristic curves disability by subgroup at follow-up. Disability-scores versus STarT-MSK_G total score; RMDQ: Roland Morris Disability Questionnaire; NDI: Neck Disability Index; SPADI: Shoulder Pain and Disability Index, subscale disability; WOMAC: Western Ontario and McMaster Universities Osteoarthritis Index, subscale disability; t2: follow-up.

(TIF)

Acknowledgments

The authors gratefully thank the participating therapeutic centres AktivioMed, Karlsruhe; Physiomed, Trier; Reha am Bahnhof-Zentrum, Neckarsulm; Reha Rondell, Brackenheim; Respoaktiv Gesundheitszentrum, Göppingen; Theraktiv, Heidelberg and Therapiezentrum Heidelberg, Heidelberg and their staff—specifically Tim Bumb, Tobias Horel and Maike Küstner—for recruitment of patients. Moreover, thanks are due to the therapeutic heads Volker Sutor, Matthias Hoppe and David Mordtan for their input on recruitment strategies, to Alexander Beckmann for his participation in cross-cultural adaptation of the STarT-MSK and his assistance on the preparation of the online questionnaires and to Mishael Adje for proreading of the manuscript.

Author Contributions

Conceptualization: Sven Karstens, Jochen Zebisch, Johannes Wey, Roger Hilfiker, Jonathan C. Hill.

Data curation: Sven Karstens.

Formal analysis: Sven Karstens, Jochen Zebisch, Roger Hilfiker, Jonathan C. Hill.

Investigation: Sven Karstens.

Methodology: Sven Karstens, Jochen Zebisch, Johannes Wey, Roger Hilfiker, Jonathan C. Hill.

Project administration: Sven Karstens, Jochen Zebisch, Johannes Wey.

Writing – original draft: Sven Karstens.

Writing – review & editing: Jochen Zebisch, Johannes Wey, Roger Hilfiker, Jonathan C. Hill.

References

1. Hoy D, March L, Woolf A, Blyth F, Brooks P, Smith E, et al. The global burden of neck pain: estimates from the global burden of disease 2010 study. *Ann Rheum Dis*. 2014; 73(7):1309–15. Epub 2014/02/01. <https://doi.org/10.1136/annrheumdis-2013-204431> PMID: 24482302
2. Cross M, Smith E, Hoy D, Carmona L, Wolfe F, Vos T, et al. The global burden of rheumatoid arthritis: estimates from the global burden of disease 2010 study. *Ann Rheum Dis*. 2014; 73(7):1316–22. Epub 2014/02/20. <https://doi.org/10.1136/annrheumdis-2013-204627> PMID: 24550173
3. Global Burden of Disease Study 2013 Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet*. 2015; 386(9995):743–800. Epub 2015/06/13. [https://doi.org/10.1016/S0140-6736\(15\)60692-4](https://doi.org/10.1016/S0140-6736(15)60692-4) PMID: 26063472
4. Hurwitz EL, Randhawa K, Yu H, Côté P, Haldeman S. The Global Spine Care Initiative: a summary of the global burden of low back and neck pain studies. *Eur Spine J*. 2018; 27(Suppl 6):796–801. Epub 2018/02/27. <https://doi.org/10.1007/s00586-017-5432-9> PMID: 29480409
5. Baxter GD, Chapple C, Ellis R, Hill J, Liu L, Mani R, et al. Six things you need to know about low back pain. *J Prim Health Care*. 2020; 12(3):195–8. <https://doi.org/10.1071/HC19117> PMID: 32988440
6. Knoop J, van Lankveld W, Geerdink FJB, Soer R, Staal JB. Use and perceived added value of patient-reported measurement instruments by physiotherapists treating acute low back pain: a survey study among Dutch physiotherapists. *BMC Musculoskelet Disord*. 2020; 21(1):120. <https://doi.org/10.1186/s12891-020-3132-9> PMID: 32093706
7. Hemingway H, Croft P, Perel P, Hayden JA, Abrams K, Timmis A, et al. Prognosis research strategy (PROGRESS) 1: a framework for researching clinical outcomes. *BMJ*. 2013; 346:e5595. Epub 2013/02/07. <https://doi.org/10.1136/bmj.e5595> PMID: 23386360
8. Hill JC, Dunn KM, Lewis M, Mullis R, Main CJ, Foster NE, et al. A primary care back pain screening tool: identifying patient subgroups for initial treatment. *Arthritis Rheum*. 2008; 59(5):632–41. Epub 2008/04/29. <https://doi.org/10.1002/art.23563> PMID: 18438893
9. Hill JC, Whitehurst DG, Lewis M, Bryan S, Dunn KM, Foster NE, et al. Comparison of stratified primary care management for low back pain with current best practice (STarT Back): a randomised controlled trial. *Lancet*. 2011; 378(9802):1560–71. Epub 2011/10/04. [https://doi.org/10.1016/S0140-6736\(11\)60937-9](https://doi.org/10.1016/S0140-6736(11)60937-9) PMID: 21963002
10. Karstens S, Krug K, Raspe H, Wunderlich M, Hochheim M, Joos S, et al. Prognostic ability of the German version of the STarT Back tool: analysis of 12-month follow-up data from a randomized controlled trial. *BMC Musculoskelet Disord*. 2019; 20(1):94. <https://doi.org/10.1186/s12891-019-2467-6> PMID: 30819162
11. Foster NE, Mullis R, Hill JC, Lewis M, Whitehurst DG, Doyle C, et al. Effect of stratified care for low back pain in family practice (IMPACT Back): a prospective population-based sequential comparison. *Ann Fam Med*. 2014; 12(2):102–11. <https://doi.org/10.1370/afm.1625> PMID: 24615305
12. Sowden G, Hill JC, Morso L, Louw Q, Foster NE. Advancing practice for back pain through stratified care (STarT Back). *Braz J Phys Ther*. 2018; 22(4):255–64. Epub 2018/07/05. <https://doi.org/10.1016/j.bjpt.2018.06.003> PMID: 29970301
13. Hsu C, Evers S, Balderson BH, Sherman KJ, Foster NE, Estlin K, et al. Adaptation and Implementation of the STarT Back Risk Stratification Strategy in a US Health Care Organization: A Process Evaluation. *Pain Med*. 2019; 20(6):1105–19. Epub 2018/10/03. <https://doi.org/10.1093/pm/pty170> PMID: 30272177
14. Campbell P, Hill JC, Protheroe J, Afolabi EK, Lewis M, Beardmore R, et al. Keele Aches and Pains Study protocol: validity, acceptability, and feasibility of the Keele STarT MSK tool for subgrouping musculoskeletal patients in primary care. *J Pain Res*. 2016; 9:807–18. Epub 2016/10/30. <https://doi.org/10.2147/JPR.S116614> PMID: 27789972
15. Dunn KM, Campbell P, Lewis M, Hill JC, van der Windt DA, Afolabi E, et al. Refinement and validation of a tool for stratifying patients with musculoskeletal pain. *Eur J Pain*. 2021; 25(10):2081–93. Epub 2021/06/09. <https://doi.org/10.1002/ejp.1821> PMID: 34101299
16. Burgess R, Mansell G, Bishop A, Lewis M, Hill J. Predictors of Functional Outcome in Musculoskeletal Healthcare: An Umbrella Review. *Eur J Pain*. 2019; 24(1):51–70. Epub 2019/09/12. <https://doi.org/10.1002/ejp.1483> PMID: 31509625
17. van den Broek AG, Kloek CJJ, Pisters MF, Veenhof C. Validity and reliability of the Dutch STarT MSK tool in patients with musculoskeletal pain in primary care physiotherapy. *PLoS One*. 2021; 16(3): e0248616. <https://doi.org/10.1371/journal.pone.0248616> PMID: 33735303

18. Dunn KM, Campbell P, Afolabi EK, Lewis M, van der Windt D, Hill JC, et al. Refinement and Validation of the Keele STarT MSK Tool for Musculoskeletal Pain in Primary Care. *Rheumatology*. 2017; 56 (suppl_2). <https://doi.org/10.1093/rheumatology/kex062.177>
19. Beaudart C, Criscenzo L, Demoulin C, Bornheim S, van Beveren J, Kaux J-F. French translation and validation of the Keele STarT MSK Tool. *European Rehabilitation Journal*. 2021; 1(1):1–7.
20. Rysstad T, Grotle M, Aasdahl L, Hill JC, Dunn KM, Tingulstad A, et al. Stratifying workers on sick leave due to musculoskeletal pain: translation, cross-cultural adaptation and construct validity of the Norwegian Keele STarT MSK tool. *Scandinavian Journal of Pain*. 2022. <https://doi.org/10.1515/sjpain-2021-0144> PMID: 35148473
21. Ben Ami N, Hill J, Pincus T. STarT MSK tool: Translation, adaptation and validation in Hebrew. *Musculoskeletal care*. 2021;n/a(n/a). <https://doi.org/10.1002/msc.1607> PMID: 34862708
22. Hay EM, Dunn KM, Hill JC, Lewis M, Mason EE, Konstantinou K, et al. A randomised clinical trial of subgrouping and targeted treatment for low back pain compared with best current care. The STarT Back Trial Study Protocol. *BMC Musculoskeletal Disord*. 2008; 9:58. Epub 2008/04/24. <https://doi.org/10.1186/1471-2474-9-58> PMID: 18430242
23. Hingorani AD, Windt DA, Riley RD, Abrams K, Moons KG, Steyerberg EW, et al. Prognosis research strategy (PROGRESS) 4: stratified medicine research. *BMJ*. 2013; 346:e5793. Epub 2013/02/07. <https://doi.org/10.1136/bmj.e5793> PMID: 23386361
24. Protheroe J, Saunders B, Bartlam B, Dunn KM, Cooper V, Campbell P, et al. Matching treatment options for risk sub-groups in musculoskeletal pain: a consensus groups study. *BMC Musculoskeletal Disord*. 2019; 20(1):271. Epub 2019/06/04. <https://doi.org/10.1186/s12891-019-2587-z> PMID: 31153364
25. Sowden G, Hill JC, Konstantinou K, Khanna M, Main CJ, Salmon P, et al. Targeted treatment in primary care for low back pain: the treatment system and clinical training programmes used in the IMPaCT Back study (ISRCTN 55174281). *Fam Pract*. 2012; 29(1):50–62. <https://doi.org/10.1093/fampra/cmr037> PMID: 21708984
26. Saunders B, Hill JC, Foster NE, Cooper V, Protheroe J, Chudyk A, et al. Stratified primary care versus non-stratified care for musculoskeletal pain: qualitative findings from the STarT MSK feasibility and pilot cluster randomized controlled trial. *BMC Fam Pract*. 2020; 21(1):31. <https://doi.org/10.1186/s12875-020-1098-1> PMID: 32046656
27. Corp N, Mansell G, Stynes S, Wynne-Jones G, Morsø L, Hill JC, et al. Evidence-based treatment recommendations for neck and low back pain across Europe: A systematic review of guidelines. *European Journal of Pain*. 2021; 25(2):275–95. <https://doi.org/10.1002/ejp.1679> PMID: 33064878
28. <Back-UP. The Back-UP web app demonstration for clinicians 2020. Available from: <https://www.youtube.com/watch?v=kgyodFFHEJ8>.
29. SoSci Survey GmbH. SoSci Survey—the Solution for Professional Online Questionnaires o.J. [024.03.2021]. Available from: <https://www.soscisurvey.de/en/index>.
30. Mookink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. COSMIN checklist manual COSMIN initiative; 2012. Available from: <http://www.cosmin.nl/images/upload/files/COSMIN%20checklist%20manual%20v9.pdf>.
31. Beaton D, Bombardier C, Guillemin F, Ferraz MB. Recommendations for the Cross-Cultural Adaptation of the DASH & QuickDASH Outcome Measures: Institute for Work & Health; 2007 [04.03.2022]. Available from: http://dash.iwh.on.ca/sites/dash/files/downloads/cross_cultural_adaptation_2007.pdf.
32. Aebischer B, Hill JC, Hilfiker R, Karstens S. German translation and cross-cultural adaptation of the STarT Back Screening Tool. *PLoS One*. 2015; 10(7):e0132068. <https://doi.org/10.1371/journal.pone.0132068> PMID: 26161669
33. Mahler C, Rochon J, Karstens S, Szecsenyi J, Hermann K. Internal consistency of the readiness for interprofessional learning scale in German health care students and professionals. *BMC Med Educ*. 2014; 14(1):145. Epub 2014/07/17. <https://doi.org/10.1186/1472-6920-14-145> PMID: 25027384
34. Guss CD. What Is Going Through Your Mind? Thinking Aloud as a Method in Cross-Cultural Psychology. *Front Psychol*. 2018; 9:1292. Epub 2018/08/29. <https://doi.org/10.3389/fpsyg.2018.01292> PMID: 30150948
35. Schmidt CO, Kohlmann T, Pflingsten M, Lindena G, Marnitz U, Pfeifer K, et al. Construct and predictive validity of the German Orebro questionnaire short form for psychosocial risk factor screening of patients with low back pain. *Eur Spine J*. 2016; 25(1):325–32. Epub 2015/08/28. <https://doi.org/10.1007/s00586-015-4196-3> PMID: 26310842
36. Stadler M, Sailer M, Fischer F. Knowledge as a formative construct: A good alpha is not always better. *New Ideas Psychol*. 2021; 60:100832. <https://doi.org/doi.org/10.1016/j.newideapsych.2020.100832>

37. Cramer H, Lauche R, Langhorst J, Dobos GJ, Michalsen A. Validation of the German version of the Neck Disability Index (NDI). *BMC Musculoskelet Disord*. 2014; 15:91. Epub 2014/03/20. <https://doi.org/10.1186/1471-2474-15-91> PMID: 24642209
38. Angst F, Goldhahn J, Pap G, Mannion AF, Roach KE, Siebertz D, et al. Cross-cultural adaptation, reliability and validity of the German Shoulder Pain and Disability Index (SPADI). *Rheumatology (Oxford)*. 2007; 46(1):87–92. Epub 2006/05/25. <https://doi.org/10.1093/rheumatology/kel040> PMID: 16720638
39. Exner V, Keel P. [Measuring disability of patients with low-back pain—validation of a German version of the Roland & Morris disability questionnaire]. *Schmerz*. 2000; 14(6):392–400. Epub 2003/06/12. <https://doi.org/10.1007/s00482000010> PMID: 12800012
40. Stucki G, Meier D, Stucki S, Michel BA, Tyndall AG, Dick W, et al. [Evaluation of a German version of WOMAC (Western Ontario and McMaster Universities) Arthrosis Index]. *Z Rheumatol*. 1996; 55(1):40–9. Epub 1996/01/01. PMID: 8868149
41. Jensen MP, Turner LR, Turner JA, Romano JM. The use of multiple-item scales for pain intensity measurement in chronic pain patients. *Pain*. 1996; 67(1):35–40. [https://doi.org/10.1016/0304-3959\(96\)03078-3](https://doi.org/10.1016/0304-3959(96)03078-3) PMID: 8895229
42. Sim J, Waterfield J. Validity, reliability and responsiveness in the assessment of pain. *Physiother Theory Pract*. 1997; 13:23–37.
43. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med*. 2016; 15(2):155–63. Epub 03/31. <https://doi.org/10.1016/j.jcm.2016.02.012> PMID: 27330520
44. Cohen J. *Statistical power analysis for the behavioural sciences*. Hillsdale, NJ: L. Erlbaum Associates; 1998.
45. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007; 60(1):34–42. Epub 2006/12/13. <https://doi.org/10.1016/j.jclinepi.2006.03.012> PMID: 17161752
46. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. 1982; 143(1):29–36. Epub 1982/04/01. <https://doi.org/10.1148/radiology.143.1.7063747> PMID: 7063747
47. Stratford PW, Riddle DL. A Roland Morris Disability Questionnaire Target Value to Distinguish between Functional and Dysfunctional States in People with Low Back Pain. *Physiotherapy Canada Physiotherapie Canada*. 2016; 68(1):29–35. <https://doi.org/10.3138/ptc.2014-85> PMID: 27504045
48. MacDermid JC, Walton DM, Avery S, Blanchard A, Etruw E, McAlpine C, et al. Measurement properties of the neck disability index: a systematic review. *J Orthop Sports Phys Ther*. 2009; 39(5):400–17. Epub 2009/06/13. <https://doi.org/10.2519/jospt.2009.2930> PMID: 19521015
49. Bieleman HJ, Reneman MF, van Ittersum MW, van der Schans CP, Groothoff JW, Oosterveld FG. Self-reported functional status as predictor of observed functional capacity in subjects with early osteoarthritis of the hip and knee: a diagnostic study in the CHECK cohort. *Journal of occupational rehabilitation*. 2009; 19(4):345–53. Epub 2009/06/27. <https://doi.org/10.1007/s10926-009-9189-9> PMID: 19557505
50. Tran G, Dube B, Kingsbury S, Tennant A, Conaghan P, Hensor E. Investigating the Patient Acceptable Symptom State cut-offs: longitudinal data from a community cohort using the Shoulder Pain and Disability Index. *Rheumatol Int*. 2019; 40. <https://doi.org/10.1007/s00296-019-04486-3> PMID: 31797040
51. Hosmer DW, Lemeshow S. *Applied logistic regression*. 2. ed. New York: Wiley; 2000.
52. Coste J, Lefrançois G, Guillemin F, Pouchot J, Rheumatology FtFSGfQoLi. Prognosis and quality of life in patients with acute low back pain: Insights from a comprehensive inception cohort study. *Arthritis Care Res (Hoboken)*. 2004; 51(2):168–76. <https://doi.org/10.1002/art.20235>
53. Woo A, Lechner B, Fu T, Wong CS, Chiu N, Lam H, et al. Cut points for mild, moderate, and severe pain among cancer and non-cancer patients: a literature review. *Annals of Palliative Medicine*. 2015; 4(4):176–83. <https://doi.org/10.3978/j.issn.2224-5820.2015.09.04> PMID: 26541396
54. Harrell FE. *rms: Regression Modeling Strategies*. R package version 5.1–2. 2018 [04.03.2022]. Available from: <https://CRAN.R-project.org/package=rms>.
55. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW. Calibration: the Achilles heel of predictive analytics. *BMC Med*. 2019; 17(1):230. Epub 2019/12/18. <https://doi.org/10.1186/s12916-019-1466-7> PMID: 31842878
56. R Development Core Team. *R: A language and environment for statistical computing* Vienna, Austria: R Foundation for Statistical Computing; 2014 [04.03.2022]. Available from: <http://www.R-project.org>.
57. Beneciuk JM, Fritz JM, George SZ. The STarT Back Screening Tool for prediction of 6-month clinical outcomes: relevance of change patterns in outpatient physical therapy settings. *J Orthop Sports Phys Ther*. 2014; 44(9):656–64. Epub 2014/08/08. <https://doi.org/10.2519/jospt.2014.5178> PMID: 25098194

58. Unal I. Defining an Optimal Cut-Point Value in ROC Analysis: An Alternative Approach. *Comput Math Methods Med*. 2017; 2017:3762651. <https://doi.org/10.1155/2017/3762651> PMID: 28642804
59. AAOS. American Academy of Orthopaedic Surgeons: Principles for Musculoskeletal Based Patient Reported Outcome-Performance Measurement Development 2018 [04.03.2022]. Available from: <https://www.aaos.org/contentassets/1cd7f41417ec4dd4b5c4c48532183b96/1188-principles-for-musculoskeletal-based-patient-reported-outcome-performance-measurement-development.pdf>.
60. Karstens S, Christiansen DH, Brinkmann M, Hahm M, McCray G, Hill JC, et al. German translation, cross-cultural adaptation and validation of the Musculoskeletal Health Questionnaire: cohort study. *Eur J Phys Rehabil Med*. 2020; 56(6):771–9. Epub 2020/09/26. <https://doi.org/10.23736/S1973-9087.20.06054-2> PMID: 32975396
61. Cella D, Hahn EA, Jensen SE, Butt Z, Nowinski CJ, Rothrock N, et al. Types of Patient-Reported Outcomes. 2015. In: *Patient-Reported Outcomes in Performance Measurement* [Internet]. Research Triangle Park (NC): RTI Press. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK424381/>.
62. Janssens A, Rogers M, Thompson Coon J, Allen K, Green C, Jenkinson C, et al. A Systematic Review of Generic Multidimensional Patient-Reported Outcome Measures for Children, Part II: Evaluation of Psychometric Performance of English-Language Versions in a General Population. *Value Health*. 2015; 18(2):334–45. <https://doi.org/10.1016/j.jval.2015.01.004> PMID: 25773569
63. Protheroe J, Saunders B, Hill JC, Chudyk A, Foster NE, Bartlam B, et al. Integrating clinician support with intervention design as part of a programme testing stratified care for musculoskeletal pain in general practice. *BMC Fam Pract*. 2021; 22(1):161. <https://doi.org/10.1186/s12875-021-01507-8> PMID: 34311697
64. Traeger AC, Hübscher M, McAuley JH. Understanding the usefulness of prognostic models in clinical decision-making. *J Physiother*. 2017; 63(2):121–5. <https://doi.org/10.1016/j.jphys.2017.01.003> PMID: 28342681
65. Main CJ, George SZ. Psychologically informed practice for management of low back pain: future directions in practice and research. *Phys Ther*. 2011; 91(5):820–4. Epub 2011/04/01. <https://doi.org/10.2522/ptj.20110060> PMID: 21451091
66. Karstens S, Kuithan P, Joos S, Hill JC, Wensing M, Steinhäuser J, et al. Physiotherapists' views of implementing a stratified treatment approach for patients with low back pain in Germany: a qualitative study. *BMC Health Serv Res*. 2018; 18(1):214. <https://doi.org/10.1186/s12913-018-2991-3> PMID: 29592802
67. Karstens S, Joos S, Hill JC, Krug K, Szecsenyi J, Steinhäuser J. General practitioners views of implementing a stratified treatment approach for low back pain in Germany: A qualitative study. *PLoS One*. 2015; 10(8):e0136119. Epub 2015/09/01. <https://doi.org/10.1371/journal.pone.0136119> PMID: 26322985
68. Karstens S, Krug K, Hill JC, Stock C, Steinhäuser J, Szecsenyi J, et al. Validation of the German version of the STarT-Back Tool (STarT-G): a cohort study with patients from primary care practices. *BMC Musculoskelet Disord*. 2015; 16(1):346. Epub 2015/11/13. <https://doi.org/10.1186/s12891-015-0806-9> PMID: 26559635
69. Furlan AD, Malmivaara A, Chou R, Maher CG, Deyo RA, Schoene M, et al. 2015 updated method guideline for systematic reviews in the Cochrane Back and Neck Group. *Spine (Phila Pa 1976)*. 2015; 40(21):1660–73. Epub 2015/07/25. <https://doi.org/10.1097/BRS.0000000000001061> PMID: 26208232
70. Remillard ML, Mazor KM, Cutrona SL, Gurwitz JH, Tjia J. Systematic review of the use of online questionnaires of older adults. *J Am Geriatr Soc*. 2014; 62(4):696–705. Epub 03/17. <https://doi.org/10.1111/jgs.12747> PMID: 24635138
71. Mokkink LB, Prinsen CA, Bouter LM, Vet HC, Terwee CB. The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) and how to select an outcome measurement instrument. *Braz J Phys Ther*. 2016; 20(2):105–13. Epub 2016/01/21. <https://doi.org/10.1590/bjpt-rbf.2014.0143> PMID: 26786084
72. Hill JC, Garvin S, Chen Y, Cooper V, Wathall S, Saunders B, et al. Stratified primary care versus non-stratified care for musculoskeletal pain: findings from the STarT MSK feasibility and pilot cluster randomized controlled trial. *BMC Fam Pract*. 2020; 21(1):30. <https://doi.org/10.1186/s12875-019-1074-9> PMID: 32046647