**Original article**

**German translation, cross-cultural adaptation and validation of the Musculoskeletal Health Questionnaire. A cohort study**

Sven KARSTENS\*1, David Hoyrup CHRISTIANSEN2,3, Melanie BRINKMANN1, Magali HAHM1, Gareth McCRAY4, Jonathan C HILL4, Stefanie JOOS5

1 Therapeutic Sciences, Department of Computer Science, Trier University of Applied Sciences, Trier, Germany

2 Occupational Medicine, Danish Ramazzini Centre, Regional Hospital West Jutland - University Research Clinic, Herning, Denmark

3 Department of Clinical Medicine, Health, Aarhus University, Denmark

4 School of Primary, Community and Social Care, Keele University, Staffordshire, United Kingdom

5 Department of General Practice, University of Tuebingen, Tuebingen, Germany

\*Corresponding author:

Sven Karstens, Therapeutic Sciences, Department of Computer Science, Schneidershof, Trier University of Applied Sciences, 54293 Trier, Germany

ORCID 0000-0001-7403-3452

mailto karstens@hochschule‐trier.de

**Abstract (291 words/max. 300)**

**BACKGROUND:** The Musculoskeletal Health Questionnaire (MSK-HQ) was developed to measure the health status of patients with various musculoskeletal conditions across multiple settings including rehabilitation.

**AIM:** Formaltranslationand cross-cultural adaptation of the MSK-HQ into German (MSK-HQG), to determine test-retest-reliability, standard error of measurement (SEM), smallest detectable change (SDC), construct validity, responsiveness, minimal important change (MIC), and to test for floor or ceiling effects.

**DESIGN:** Cohort study with six weeks follow-up.

**SETTING:** Seven physiotherapy clinics/rehabilitation centres.

**POPULATION:** Patients with a referral for physiotherapy indicating musculoskeletal complaints of the spine or extremities.

**METHODS:** Translation and cross-cultural adaptation were carried out in accordance with guidelines provided by the developers. As reference standards we used pain intensity (0-10 numeric rating scale), quality of life (EQ5D-5L) and disability measures (RMDQ, NDI, WOMAC and SPADI) that were combined using z-scores.

**RESULTS:** On 100 patients (age 44.8±13.4 years, 66% female) the test-retest-reliability intraclass correlation coefficient was 0.87 (95% CI 0.72; 0.93) and for construct validity correlation with the combined disability measure was rs = -0.81 (95% CI -0.88, -0.72), the SEM was 3.4, the SDC (individual) 9.4, and the MIC 8.5.

**CONCLUSIONS:** Overall, the study provides evidence for good reliability and validity for the MSK-HQG. Further studies in different settings and diagnostic subgroups should follow to better understand the psychometric properties of this measure in primary care, rehabilitation and specialist care settings.

**CLINICAL REHABILITATION IMPACT:** The results demonstrate that the MSK-HQG has sufficient psychometric properties for use in musculoskeletal research and practice. However, the SDC should be kept in mind when using the tool for individual patients. The MSK-HQG has the advantage of being a single instrument that can measure musculoskeletal health status across different pain sites, reducing the burden from the use of multiple tools.

**Keywords:** Patient Reported Outcomes, Psychometrics, Musculoskeletal Diseases, Responsiveness

Introduction

Musculoskeletal (MSK) diseases, such as back pain, neck pain or osteoarthritis are among the leading causes of disability-adjusted life years and often have a chronic course.1-3 The prevalence of and disability caused by some MSK conditions has risen substantially over the past decades and it is assumed that this development continues due to aging populations.4 Therefore, prevention and rehabilitation for these conditions is a research priority.4, 5

Patient-reported outcomes (PRO) are used to gather information about the state of health from the patients themselves, most often using questionnaires. Over the recent years PROs have increasingly been adopted in research and clinical practice.6, 7 Domains addressed for patients with musculoskeletal complaints often include pain, disability and quality of life.8 Different factors influence the course of musculoskeletal conditions and are important to consider.9 PROs can largely be divided into two categories: disease-specific measures and generic measures which can be used across different diagnoses.10 Established examples are the back pain specific Roland Morris Disability Questionnaire (RMDQ) or the generic EuroQol five-dimension, five-level version (EQ-5D-5L).11 An instrument used in rehabilitation and specifically developed for patients with musculoskeletal conditions, covering the spine and the extremities is the Short Musculoskeletal Function Assessment (SMFA). Such cross cutting instruments are advantageous for clinical practice meaning a reduced administrative burden especially as often patients present with MSK pain from several sites simultaneously.8 Nevertheless, the SMFA has some substantial limitations as it is lengthy with 46 items and its development and psychometric properties were primarily undertaken on patients with trauma of an extremity rather than common musculoskeletal complaints.12

Recently the Musculoskeletal Health Questionnaire (MSK-HQ) was developed and validated.13 With 14 items it is a short and therefore practical instrument and is intended to be used throughout the musculoskeletal care pathway and for different pain sites (e.g. back or knee). Initial validation of the MSK-HQ was promising as it has demonstrated high completion rates, good test-retest reliability and strong convergent validity, when compared to relevant condition specific instruments.13 Patients and clinicians were involved in the development process to facilitate relevant content validity and ensure its usefulness for routine clinical practice, e.g. for patient communication and clinical decision making. With items on emotion, sleep and self-confidence to manage the condition, alongside traditional items on pain and disability, it provides a holistic view of the impact of the condition fitting to the rehabilitation perspective.13, 14

The aim of this study was, to translate and cross-culturally adapt the MSK-HQ into German and evaluate its measurement properties in terms of 1) test-retest-reliability, 2) standard error of measurement (SEM) and smallest detectable change (SDC), 3) construct validity, 4) responsiveness and 5) minimal important change (MIC).

Materials and methods

Translation and cross-cultural adaptation

The translation and cross-cultural adaptation were performed in collaboration with a representative of the research group developing the questionnaire (JCH). It was carried out in accordance with guidelines provided by the licence owner that are closely aligned to internationally accepted standards.15 Two native German speaking translators, one of whom was a professional translator with experience in the medical field and one a lay person without medical knowledge. To consider language variations across different German-speaking countries, two additional translators from Switzerland and Luxembourg (native speakers) participated. The translators were asked to note any remarks and questions arising during the process. The four forward translations were synthesized by a German member of the steering committee into a final forward version and adjusted in correspondence with the translators. A concept elaboration document was provided by the developers to support the process. Within the document word meanings were clarified and specific constructs were explained. The backward translations were done by two English native speakers, one with medical knowledge. The translations were sent for discussion to the developers of the original English version, who confirmed they were appropriate. During the process difficult formulations were noted and different suggestions documented to test alternatives during pre-tests.

To check for acceptability and comprehensibility a pre-test was carried out with 14 German patients from a physiotherapy clinic. Eligibility criteria for participation were age ≥18 and a physiotherapy referral due to musculoskeletal complaints. Participants were selected using a purposeful sampling strategy considering the criteria age, gender, region of complaints, duration of symptoms, and number of treatment sessions. Utilized methods were the Think-Aloud and Response Latency method and a semi-standardized interview guide considering experiences from the translation process.16, 17 The discussions were audiotaped, transcribed and coded using a content analytic approach.18 Pauses for thoughts were noted as remarkable, if they lasted longer than five seconds. Moreover, it was documented if the patients did not answer questions of the German version of the MSK-HQ (MSK-HQG) if they were not reminded.

Study design for psychometric properties

This was a prospective cohort study. Patients were recruited in seven facilities (1 hospital outpatient physiotherapy clinic, 2 outpatient rehabilitation centres and 4 physiotherapeutic clinics) spread across three federal states. It took place between November 2017 and December 2018.

To be included the patients had to be 18 to 70 years and visiting the clinic with a referral for physiotherapy indicating musculoskeletal complaints of the spine or extremities. The nature of the complaints could be degenerative or traumatic and concern the shoulder, the knee or hip, the cervical or the lumbar spine. Patients who had undergone surgery in one of these regions were also included. Patients with a private health insurance (approximately 10% of the potential sample19) were not invited for participation since their referral forms did not give sufficient information to check inclusion criteria. Excluded were patients not able to read and understand German questionnaires and those who had undergone physiotherapy treatment within the previous 12 weeks.

Patients were asked to complete questionnaires at three time points: initial assessment (T0, preferably before contact with the therapist), retest (T1, before second treatment session) and follow-up (T2, questionnaire sent five weeks after T0). If it was necessary due to organizational structures in the physiotherapy clinics, instead of filling in the questionnaires before the first contact with the therapist, it was possible to answer them immediately afterwards. The concrete time point (before or after first treatment session) was documented for the individual patient on a pre-defined form.

All participants gave written informed consent prior to participation. Ethical approval for the trial was granted by the Ethics Committee of the Faculty of Medicine of the University of Tübingen (No 699/2016BO2, Nov 17 2016). For non-consenters age and gender were documented.

Instruments

All patients were given the MSK-HQG. The instrument consists of 14 items. With two exceptions the responses of each item are operationalized on a five-point verbal rating scale ranging from 4 = ‘not at all’ to 0 = ‘extremely’. Similar to the original English version, the responses of items 12 and 13 (‘understanding of your condition’, item 12; ‘confidence […] to manage your symptoms’, item 13) are given in reverse order, since in their case, in contrast to the other items, the answer ‘not at all’ indicates a deficit with 0 points being allocated. The points were combined into a sum score ranging from 0 to 56 points, with a higher score suggesting a better musculoskeletal health status.13 To measure quality of life and to enable comparability with the original version the EQ-5D-5L was used.13, 20 Three eleven-point box-scales for least, average (over the previous two weeks), and current pain were used to measure pain intensity.21 Current health status was captured using a five-point verbal rating scale ranging from overall poor to excellent health (a presentation of instruments per time point is given in **Table I**).

To assess disability available and validated German questionnaires were chosen. Depending on the patients’ complaints one of the following instruments were added: German version of the Neck Disability Index (NDI)22, Shoulder Pain and Disability Index (SPADI)23, Roland Morris Disability Questionnaire (RMDQ)24 or Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)25.

In addition to the MSK-HQG at T1, the patients had to answer a question concerning global perceived change in comparison to the time point of the initial assessment using a five-point verbal rating scale with the following response options ‘complaints much worse’, ‘complaints slightly worse’, ‘complaints not changed’, ‘complaints much better’ or ‘complaints much better’.26, 27 At T2 in addition to the MSK-HQG the patients again received the EQ-5D-5L11, the questions on pain intensity and the scale on global perceived change. Again, at T2 all participants completed the complaint specific reference standard measure.

Statistical analyses

Recommendations for sample-size estimates for quantitative studies on health status questionnaires range between 50 and 100 participants (28, 31). To secure a sufficient size and in case of an advantageous distribution of subgroups, to enable preliminary subgroup analyses, a sample of 130 patients was aimed for.

To examine *test-retest reliability* the intraclass correlation coefficient (ICC (2,1), based on two-way random effect, absolute agreement model) was used to test overall MSK-HQG score agreement between T0 and T1.28, 29 An ICC above 0.70 is considered adequate.29 Additionally, Cohen’s quadratic weighted Kappa coefficient (κW) was calculated to examine individual item agreement.30

The *standard error of measurement* (SEM=SDDiff/√2) and *smallest detectable change* (SDC=1.96×√2×SEM) were calculated to determine the extent to which changes can be considered relevant to patients or differences can be assumed being no result of measurement error.28, 29, 31

To test for *internal consistency* Cronbach’s Alpha was determined. A coefficient above 0.7 is considered to be acceptable, above 0.8 to be good and above 0.9 to be excellent.27

To assess *construct validity*, Spearmans Rho correlation coefficients were calculated. Spearmans Rho was chosen since it does not require linearity or interval scaling.32 The following combinations were analysed: MSK-HQG/pain, MSK-HQG/quality of life (EQ-5D-5L) and MSK-HQG/disability (combined reference standard).. Since the sample size clearly fell short of the recommended threshold (≥ 50 participants29) for each pain site within the timeframe set for recruitment, the scores of the reference standard measures for the patients with back, neck, shoulder and lower extremity complaints (RMDQ, NDI, WOMAC disability-scale and SPADI disability-scale) were combined in one variable after standardization (z-scores).33 Since combination of scores is not a standard procedure, a sensitivity analysis for MSK-HQG/disability was done by calculation of four correlations, each time excluding one of the subgroups (back, neck, shoulder and lower extremity) from the combined score.

A priori, it was hypothesised that the MSK-HQG would correlate more strongly (higher values) with the reference standard measure scores than with the EQ-5D-5L scores 28. In addition change scores were compared for the named combinations.

The effect size (ES = mean change/SDinitial) for the change at T2 was calculated for the MSK-HQG and to enable comparison the same coefficient was calculated for the EQ-5D-5L27. Moreover, to support interpretation Spearmans Rho coefficients were calculated comparing the MSK-HQG and EQ-5D-5L change scores (with values for global perceived change).

For *interpretability* the minimal important change (MIC) values were estimated based on Receiver Operating Characteristic (ROC) curves using the closest to the top-left method.34 Patients were divided into two groups based on their global perceived change response with the categories ‘unchanged’ and ‘slightly better’ (group 1) dichotomised against ‘much improved’ (group 2). Responsiveness to worsening was not analysed, as only few patients rated their complaints to be ‘worse’ or ‘much worse’. As MIC values can be affected by scores from initial assessment, the analysis was repeated for relative change scores (i.e., change scores expressed as percentages of the initial scores).35 Alternative methods to derive MIC estimates do exist.36, 37 However, we chose to use the ROC curve method to preserve comparability with other studies in the field38 and communicability with readers unfamiliar with alternative methods.

*Floor and ceiling effects* were defined as present if > 15% of the responders achieved the lowest or highest possible MSK-HQ score.29

In case of missing values/loss to follow-up the cases were excluded from the respective analysis. All statistical tests were two-sided and a significance level of alpha = 5% was used.

Results

Translation and cross-cultural adaptation

Several changes were made during the forward-backward translation process. The adaptation of the term ‘symptom’ concerned several items. In agreement with the developers the term ‘complaint’ was used as reference as its translation is easier to understand by German patients. Moreover, in line with the MSK-HQ’s concept elaboration document the terms ‘condition’ (item 12) and ‘work’ (item 6) were complemented by the terms ‘medical’ and ‘occupational’ to clarify the constructs. Alternative formulations were documented during forward-/backward-translation for items 1, 2, 5, 8 and 12 (for the phrases ‘usual’ (item 1 and 2), ‘How much has it been a problem’ (item 5), ‘Needing help’ (Item 8) and ‘Understanding’ and ‘Thinking about your joint or muscle symptoms, […]?’ (Item 12) from the original version). Subsequently, they were tested during pre-tests with the 14 patients. The analysis showed that the MSK-HQG fits to the original MSK-HQ. For items 3, 6, 7, 12 and 13 pauses for thoughts of more than five seconds were noted, but only one patient skipped one item without answering it (item 12). The response options for item 12 were adapted, since in the first version the grading was perceived as inadequate. Pilot testing results were sent to and confirmed by the original developers.

Psychometric properties

Demographic data

Of the 178 people who were invited 100 (56%) agreed to participate. The mean age of the participants was 44.8 (SD 13.4) years. Consenters were more often female than non-consenters (66% versus 45%) and of younger age (mean age: 46.6 versus 53.4). Seventy-nine patients (79%) answered the questionnaires before the first treatment session, and 15 (15%) answered it afterwards (for 6 patients the time point before/after first treatment was not documented). There were 43 lower back, 28 neck, 14 shoulder and 15 hip/knee patients. Eight patients (8%) received the treatment due to previous surgery. More details on the characteristics of the study population are given in **Table II**.

Descriptive analysis of initial and follow-up scores

The follow up questionnaires sent to the patients at T1 and T2 were returned by 83% and 80% of the participants, respectively. Non-responders at T1 and T2 did not differ noticeably from responders considering gender and age. The median time interval between T0 and T1 was 4 (IQR 5) days and between T0 and T2 57 (IQR 15) days. With 0% of patients having an MSK-HQ total score of 0 points and 1% with the maximal score of 56 points no floor or ceiling effects were observed. The mean MSK-HQ score for the total sample and per pain site are presented in **Table II** and **Table III.**

Reliability

For the *test-retest reliability* (T0 to T1) of the overall MSK-HQG score the ICC was 0.87 (95% CI 0.72; 0.93. The SEM was 3.4 and the SDCind was 9.4. For individual items the median κWwas 0.75 (range 0.41 (item 1) to 0.84 (item 9)) (**Table IV**).39 Cronbach’s Alpha was α = 0.92.

Construct validity

Correlations for MSK-HQG against quality of life (EQ-5D-5L) and pain intensity measures at t0 were rs = 0.75 (95% CI 0.65, 0.83) and rs = -0.65 (95% CI -0.75, -0.51), respectively. Correlation for MSK-HQG with the standardized/combined disability score was also large at rs = -0.81 (95% CI -0.88, -0.72) and was higher than the MSK-HQG’s correlation with quality of life measured by the EQ-5D-5L. **Table V** displays the sensitivity analysis with single subgroups excluded.

Responsiveness

The correlation between the MSK-HQ change scores and change scores for EQ-5D-5L, pain intensity and combined disability change score were significant with rs = 0.64 (95% CI 0.49, 0.76), rs = -0.54. (95% CI -0.68, -0.36), rs = -0.62 (95% CI -0.75, -0.44), respectively.

According to the question on global perceived change 22 patients (27.8 %) reported to be ‘improved’, 38 (48.1 %) to be ‘much improved’, 13 patients (16.5 %) were ‘unchanged’ and 6 patients (7.6 %) were ‘worse’ at T2. The overall mean change score for the MSK-HQG T0 to T2 was 8.5 (SD 11.3). A moderate effect size of ES = 0.75 resulted. The mean change score and the effect size for EQ-5D-5L were 0.09 (SD 0.21) and ES = 0.48. The correlations between global perceived change at T2 and MSK-HQG change scores were rs = -0.53 (95% CI 0.33, 0.71), for EQ-5D-5L rs = 0.49 (95% CI 0.27, 0.67).

The ROC curves calculated for *clinical* *interpretability* are illustrated in **Figure 1a and b.** Compared are ROC curves for absolute and relative change scoresof the MSK-HQG and the EQ-5D-5L index at T2. For both cohorts the MSK-HQG demonstrated greater responsiveness (higher AUCs), than the EQ-5D-5L.

Minimally important change

The minimal important change is MICROC = 8.5 for the MSK-HQG with an area under the curve of AUC = 0.78 (95% CI 0.67, 0.89). The area for the EQ-5D-5L was AUC = 0.68 (95% CI 0.55, 0.80). The percentage of patients reaching the MIC-threshold calculated for the MSK-HQG was 48.1%. For the relative change score the MICROC = 22% with an AUC = 0.73 (95% CI 0.61, 0.85) in contrast to an area of AUC = 0.67 (95% CI 0.54, 0.79) for the EQ-5D-5L. The percentage of patients reaching the MIC-threshold for relative change scores calculated for the MSK-HQG was 45.5%.

Discussion

The results of this study demonstrate that the MSK-HQG has good test-retest-reliability and excellent internal consistency. Construct validity was confirmed and performance compared against existing site-specific disability measures, pain intensity, and a commonly used health related quality of life measure was good. Moreover, the instrument showed to have no floor or ceiling effects.

The MSK-HQ enriches the library of self-reported tools available for assessment in physiotherapy and rehabilitation settings. With a growing number of different language versions of the MSK-HQ, internationally coordinated studies to compare patients’ health outcomes in rehabilitation and across settings will be possible. It has been called for such instruments which are practical and moreover applicable for multiple musculoskeletal pain sites.8 The findings of this study are important as they add knowledge about the psychometric properties of instruments fitting these needs.

To determine construct validity, the MSK-HQG score was administered together with scores of reference instruments. Used were standard measures for disability and health related quality of life. Overall good correlation coefficients resulted. This fits to the idea that the construct ‘Musculoskeletal quality of health’ of the MSK-HQ is related to both constructs, but still is unique. The original version of the MSK-HQ showed coefficients around 0.8 for health related quality of life of patients with inflammatory arthritis and of patients from a physiotherapy cohort, which is a slightly stronger correlation than demonstrated in this study for the MSK-HQG.13, 40

Clinicians need to be aware, that the SDC (9.4) for the MSK-HQG was roughly one point higher than the determined MIC (8.5). This is important as we suggest that clinicians should only consider an individual’s MSK-HQG score, which changes by at least 10 points, to be considered clinically meaningful. The MIC for the original version of the MSK-HQ of 5.5 points, calculated by Price et al., was distinctly lower41, but next to differences between the original and the translation this might be related to differences between the samples with many patients who received surgery in the UK-cohort or a different anchoring. However, other coefficients were comparable. For example, the correlation of the MSK-HQG change scores with those of the EQ-5D-5L was moderate, which was the same for the original UK physiotherapy cohort reported by Price et al.41 Also in alignment with previous work the effect size for the MSK-HQG was higher when compared to the EQ-5D-5L, although the CIs were overlapping.41 It will be interesting to keep under review if future studies will underpin these larger effect sizes for the German version. They are a relevant factor when comparing the MSK-HQ against the EQ-5D-5L, as for the latter the limited sensitivity to change is documented.42

With pain, fatigue, physical function, symptom interference, sleep, self-efficacy and psychological well-being the MSK-HQ comprises seven domains. The items ‘understanding of your condition’ (item 12) and ‘confidence […] to manage your symptoms’ (item 13) were shown to be least stable in two previous studies on the original version.13, 40 Norton et al.40 have highlighted that ‘understanding’ or in terms of the self-regulation model of illness ‘illness coherence’43 and ‘confidence’ in the sense of ‘self-efficacy’44 are not well related to the severity of MSK-conditions. They even calculated a reduced score excluding these two items. Nevertheless, we would recommend these items are still used as they address important goals to achieve with the patient45 and emerged as important to patients when the MSK-HQ items were selected. The possibility of measuring the progress of a patient’s understanding of their condition, as well as how their self-efficacy changes with treatment, has been shown to have strong clinical and patient face validity.40 For the MSK-HQG the properties determined for the addressed items were not remarkable.

The key strengths of this study were the rigorous use of pre-defined guidelines including intensive pre-tests and a low loss to follow-up. Limitations are that study consenters were more often female and younger of age then the non-consenters and that subgroup analyses were not possible. Moreover, it has to be noted that in some cases the MSK-HQG was answered after first contact with the therapist and not before. This might have influenced the coefficients, but with only 15% of cases we consider its potential impact to have been small and for determining responsiveness it is more likely to have led to an underestimation rather than overestimation of our estimates.

The sample size reached the requirement of 50 to 100 participants recommended in the literature,28, 29 but even after an extended recruitment period the size was not sufficient to carry out preliminary subgroup analyses for each pain site (back, neck, shoulder or hip/knee). Studies with larger samples enabling such analyses or focusing on one pain site should be conducted in future. Moreover, future studies might include a head-to-head comparison with the MSK-HQG against the SMFA, which is an alternative instrument in German that is designed for musculoskeletal patients across a range of diagnoses.12 Whilst, the findings from this study suggest the reliability and responsiveness of the two instruments maybe similar, we would recommend that a future study performs a direct comparison.

The Cochrane Back and Neck Group considers that a minimum loss to follow-up of 70% for long-term and 80% for short-term time periods is unlikely to lead any retention bias. With 79% our follow-up result rose above the first threshold and gets even close to the second one for long-term time periods.46 The return for test-retest-reliability was lower and therefore the possibility for bias in this case is higher. Still, the CI for the calculated ICC seem robust.

Almost half of the participants were seeking care due to back pain. This meant that they had a strong influence on the established coefficients, likely indicated by the sensitivity analysis conducted for construct validity. Nevertheless, all combinations of the sensitivity analysis showed at least a good correlation.

Conclusions

With good test-retest-reliability, and good construct validity the MSK-HQG fulfills the demand for instruments that are applicable across multiple musculoskeletal conditions and across different rehabilitation settings, although the relatively high MIC should be kept in mind. With the growing number of different language versions available internationally coordinated studies will be enabled.

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;73:1309-15.

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;73:1316-22.

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;386:743-800.

4. Hurwitz EL, Randhawa K, Yu H, Cote 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.

5. Cross M, Smith E, Hoy D, Nolte S, Ackerman I, Fransen M, et al. The global burden of hip and knee osteoarthritis: estimates from the global burden of disease 2010 study. Ann Rheum Dis 2014;73:1323-30.

6. Holmes MM, Lewith G, Newell D, Field J, Bishop FL. The impact of patient-reported outcome measures in clinical practice for pain: a systematic review. Qual Life Res 2017;26:245-57.

7. Refolo P, Minacori R, Mele V, Sacchini D, Spagnolo AG. Patient-reported outcomes (PROs): the significance of using humanistic measures in clinical trial and clinical practice. Eur Rev Med Pharmacol Sci 2012;16:1319-23.

8. AAOS. American Academy of Orthopaedic Surgeons: Principles for Musculoskeletal Based Patient Reported Outcome-Performance Measurement Development; [cited 26.04.2020]. Available from: <https://www.aaos.org/uploadedFiles/PreProduction/About/Opinion_Statements/position/1188%20Principles%20for%20Musculoskeletal%20Based%20Patient%20Reported%20Outcome-Performance%20Measurement%20Development.pdf>.

9. Evanoff B, Dale AM, Descatha A. A conceptual model of musculoskeletal disorders for occupational health practitioners. Int J Occup Med Environ Health 2014;27:145-8.

10. Nixon A, Wild D, Muehlhausen W. Patient Reported Outcomes: An overview. Turin: SEEd Medical Publishers; 2015.

11. Brooks R. EuroQol: the current state of play. Health Policy 1996;37:53-72.

12. Bouffard J, Bertrand-Charette M, Roy JS. Psychometric properties of the Musculoskeletal Function Assessment and the Short Musculoskeletal Function Assessment: a systematic review. Clin Rehabil 2016;30:393-409.

13. Hill JC, Kang S, Benedetto E, Myers H, Blackburn S, Smith S, et al. Development and initial cohort validation of the Arthritis Research UK Musculoskeletal Health Questionnaire (MSK-HQ) for use across musculoskeletal care pathways. BMJ Open 2016;6:e012331.

14. Pfeifer K, Sudeck G, Bruggemann S, Huber G. [DGRW-update: exercise therapy in medical rehabilitation - effects, quality, perspectives]. Rehabilitation (Stuttg) 2010;49:224-36.

15. Beaton D, Bombardier C, Guillemin F, Ferraz MB. Recommendations for the Cross-Cultural Adaptation of the DASH & QuickDASH Outcome Measures: Institute for Work & Health; [cited 26.04.2020]. Available from: <http://dash.iwh.on.ca/sites/dash/files/downloads/cross_cultural_adaptation_2007.pdf>.

16. Guss CD. What Is Going Through Your Mind? Thinking Aloud as a Method in Cross-Cultural Psychology. Front Psychol 2018;9:1292.

17. Fazio R. A practical guide to the use of response latency in social psychological research. In: Hendrick C, Clark MS, editors. Review of Personality and Social Psychology. 11. London: Sage Publications; 1990. p. 74-97.

18. Mayring P. Qualitative Content Analysis; [cited 26.04.2020]. Available from: <http://www.qualitative-research.net/index.php/fqs/article/view/1089/2385>.

19. Ärzteblatt. [Nearly nine millons privatly insured]; [cited 30.07.2017]. Available from: <https://www.aerzteblatt.de/nachrichten/52395/Fast-neun-Millionen-Privatversicherte-in-Deutschland>.

20. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res 2011;20:1727-36.

21. Sim J, Waterfield J. Validity, reliability an responsiveness in the assessment of pain. Physiother Theory Pract 1997;13:23-37.

22. 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.

23. 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;46:87-92.

24. 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;14:392-400.

25. 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;55:40-9.

26. Schulte E, Berghöfer A, Hermann K. [Analysis of care pathways for patients with primary therapy resistent subacute and chronic pain]: Charité Universitätsmedizin; 2006.

27. Fayers PM, Machin D. Quality of life the assessment, analysis, and reporting of patient-reported outcomes. Chichester: John Wiley & Sons Inc; 2016.

28. 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;20:105-13.

29. 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;60:34-42.

30. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. Phys Ther 2005;85:257-68.

31. Polit DF. Getting serious about test-retest reliability: a critique of retest research and some recommendations. Qual Life Res 2014;23:1713-20.

32. Schober P, Boer C, Schwarte LA. Correlation Coefficients: Appropriate Use and Interpretation. Anesthesia & Analgesia 2018;126.

33. Weiß C. Basiswissen Medizinische Statistik. Berlin: Springer; 2019.

34. Froud R, Abel G. Using ROC curves to choose minimally important change thresholds when sensitivity and specificity are valued equally: the forgotten lesson of pythagoras. theoretical considerations and an example application of change in health status. PloS one 2014;9:e114468-e.

35. Olsen MF, Bjerre E, Hansen MD, Tendal B, Hilden J, Hrobjartsson A. Minimum clinically important differences in chronic pain vary considerably by baseline pain and methodological factors: systematic review of empirical studies. J Clin Epidemiol 2018;101:87-106.e2.

36. Terluin B, Eekhout I, Terwee CB. The anchor-based minimal important change, based on receiver operating characteristic analysis or predictive modeling, may need to be adjusted for the proportion of improved patients. J Clin Epidemiol 2017;83:90-100.

37. Terluin B, Eekhout I, Terwee CB, de Vet HC. Minimal important change (MIC) based on a predictive modeling approach was more precise than MIC based on ROC analysis. J Clin Epidemiol 2015;68:1388-96.

38. Christiansen DH, McCray G, Winding TN, Andersen JH, Nielsen KJ, Karstens S, et al. Measurement properties of the Musculoskeletal Health Questionnaire (MSK-HQ): a between country comparison. Research Square [Internet]. 2020 [cited 26.04.2020]. Available from: <http://europepmc.org/abstract/PPR/PPR151080>

<https://doi.org/10.21203/rs.2.19546/v2>

39. Altman DG. Practical statistics for medical research. London: Chapman and Hall; 1991. XII, 612 S. p.

40. Norton S, Ellis B, Santana Suárez B, Schwank S, Fitzpatrick R, Price A, et al. Validation of the Musculoskeletal Health Questionnaire in inflammatory arthritis: a psychometric evaluation. Rheumatology 2018;58:45-51.

41. Price AJ, Ogollah R, Kang S, Hay E, Barker KL, Benedetto E, et al. Determining responsiveness and meaningful changes for the Musculoskeletal Health Questionnaire (MSK-HQ) for use across musculoskeletal care pathways. BMJ Open 2019:(accepted).

42. Tordrup D, Mossman J, Kanavos P. Responsiveness of the EQ-5D to clinical change: is the patient experience adequately represented? Int J Technol Assess Health Care 2014;30:10-9.

43. Leventhal H, Brissette I, Leventhal EA. The common-sense model of self-regulation of health and illness In: Cameron LD, Leventhal H, editors. The Self-regulation of Health and Illness Behaviour: Routledge; 2003. p. 42-65.

44. Bandura A. Perceived Self-Efficacy in Cognitive Development and Functioning. Educational Psychologist 1993;28:117-48.

45. Hoffman BM, Papas RK, Chatkoff DK, Kerns RD. Meta-analysis of psychological interventions for chronic low back pain. Health Psychol 2007;26:1-9.

46. 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;40:1660-73.

Notes

Conflicts of interest

The authors declare that there is no conflict of interest.

Funding

Sven Karstens received an internal Grant from the University of Applied Sciences Trier for study-material and staff. The development of the original version of the MSK-HQ was funded by Versus Arthritis (UK Ref. 20518).

Contributions

SK: designed the study, recruited the participating clinics, organized data collection, analysed and interpreted the data and wrote the manuscript

DHC: designed the study, analysed and interpreted the data and critically revised the manuscript

MB: designed the study, recruited the participating clinics, organized data collection, and critically revised the manuscript

MH: designed the study, organized data collection and critically revised the manuscript

GMC: analysed and interpreted the data and critically revised the manuscript

JCH: designed the study, analysed and interpreted the data and revised the manuscript

SJ: designed the study, analysed and interpreted the data and revised the manuscript

Acknowledgments

The authors thank Leonhard Engeler for his work on the cross-cultural adaptation of the MSK-HQ including his work on the pre-tests. Moreover, the authors gratefully acknowledge support of the translators like Bernhard Aebischer and of the participating therapeutic centres for recruitment of the patients.

**Table I Time points and instruments administered**

|  |  |
| --- | --- |
| **Time point** | **Instruments** |
| T0 (initial assessment) | MSK-HQG, EQ-5D-5L, pain intensityDepending on diagnosis: RMDQ, NDI, SPADI, WOMAC |
| T1 (retest) | MSK-HQG, global perceived change |
| T2 (responsiveness) | MSK-HQG, EQ-5D-5L, pain intensity, global perceived changeDepending on diagnosis: RMDQ, NDI, SPADI, WOMAC |

EQ-5D-5L: EuroQol five-dimension scale, MSK-HQG: Musculoskeletal Health Questionnaire, German Version, NDI: Neck Disability Index, RMDQ: Roland Morris Disability Questionnaire, SPADI: Shoulder Pain and Disability Index, WOMAC: Western Ontario and McMaster Universities Osteoarthritis Index

**Table II Characteristics of the study population**

|  |  |  |
| --- | --- | --- |
| **Mean age in years (SD) n** | 44.8 (13.4) | N = 100 |
| **Gender %, n** |  |  |
| **Female** | 66 | 66 |
| **Male** | 34 | 34 |
|  **Total** |  | 100 |
| **Mean Body-Mass-Index in kg/m2 (SD) n** | 25.9 (4.79) | 96 |
| **Education %, n** |  |  |
| **No graduation** | 2.1 | 2 |
| **O-level** | 34.4 | 33 |
| **A-level** | 63.5 | 61 |
| **Total** |  | 96 |
| **Work status %, n** |  |  |
| **Working ≥ 35 hours** | 49.5 | 48 |
| **Working < 35 hours** | 26.8 | 26 |
| **Trainee/retrainee/prentice** | 23.7 | 23 |
| **Total** |  | 97 |
| **Sickness certificate previous 12 weeks %, n** |  |  |
| **No** | 65.3 | 62 |
| **Yes** | 34.7 | 33 |
| **Total** |  | 95 |
| **Symptom duration %, n** |  |  |
| **< 12 weeks** | 41.8 | 41 |
| **≥ 12 weeks** | 58.2 | 57 |
| **Total** |  | 98 |
| **Mean (SD) pain intensity (0-10) n** | 3.6 (1.91) | 98 |
| **Mean (SD) MSK-HQG, n** | 31.9 (11.2) | 97 |
| **Mean (SD) EQ-5D-5L, n** | 0.76 (0.186) | 98 |

Abbreviations: EQ-5D-5L: EuroQol five-dimension scale, MSK-HQG: Musculoskeletal Health Questionnaire, German Version, SD: standard deviation

**Table III MSK-HQ-G scores per diagnostic group**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Time point | T0 |  |  | T1 |  |  | T2 |  |  |
| Group | **Mean** | **SD** | **n** | **Mean** | **SD** | **n** | **Mean** | **SD** | **n** |
| Back | 30.83 | (12.90) | 42 | 30.27 | (11.62) | 30 | 39.72 | (9.30) | 32 |
| Neck | 34.86 | (8.91) | 28 | 35.65 | (8.63) | 20 | 44.29 | (8.70) | 21 |
| Hip/knee | 31.14 | (11.61) | 14 | 34.17 | (10.03) | 12 | 38.00 | (14.19) | 12 |
| Shoulder | 30.08 | (9.04) | 13 | 32.67 | (3.84) | 09 | 37.29 | (9.68) | 14 |
| Total | 31.9 | (11.20) | 97 | 32.7 | (9.95) | 71 | 40.2 | (10.23) | 79 |

Abbreviations: MSK-HQG: Musculoskeletal Health Questionnaire, German Version, SD: standard deviation, T0: initial assessment, T1: retest, T2: responsiveness. N = 100

**Table IV Quadratic weighted Kappa coefficients of single item test-retest of the MSK-HQG per item**

|  |  |  |
| --- | --- | --- |
| **Item** | **Kappa** | **Lower CI, upper CI (95%)** |
| **1** | 0.41 | 0.17, 0.62 |
| **2** | 0.76 | 0.59, 0.88 |
| **3** | 0.71 | 0.55, 0.82 |
| **4** | 0.78 | 0.68, 0.85 |
| **5** | 0.59 | 0.41, 0.75 |
| **6** | 0.76 | 0.63, 0.85 |
| **7** | 0.77 | 0.65, 0.86 |
| **8** | 0.79 | 0.64, 0.91 |
| **9** | 0.84 | 0.73, 0.92 |
| **10** | 0.84 | 0.74, 0.90 |
| **11** | 0.64 | 0.43, 0.79 |
| **12** | 0.66 | 0.42, 0.80 |
| **13** | 0.75 | 0.62, 0.83 |
| **14** | 0.66 | 0.46, 0.80 |

Abbreviations: MSK-HQG: Musculoskeletal Health Questionnaire, CI: Confidence interval

**Table V Sensitivity analysis subgroups**

|  |  |  |
| --- | --- | --- |
| **Subgroups** | **rs** | **Lower CI, upper CI (95%)** |
| Back, neck and hip/knee (1) | -0.82 | -0.888, -0.730 |
| Back, neck and shoulder (2) | -0.82 | -0.888, -0.728 |
| Back, hip/knee and shoulder (3) | -0.84 | -0.900, -0.736 |
| Neck, hip/knee and shoulder (4) | -0.75 | -0.854, -0.597 |

rs: Spearmans correlation coefficient, CI: Confidence interval, (1) patients with shoulder complaints excluded, (2) hip/knee excluded, (3) neck excluded, (4) back excluded

**Figure 1a and b Receiver-operating-characteristic curves representing (a) absolute and (b) relative T2 change scores of the MSK-HQ-G and EQ-5D-5L**

a) Blue line: MSK-HQ-G AUC = 0.78; red line: EQ-5D-5L AUC = 0.68

b) Blue line: MSK-HQ-G AUC = 0.73; red line: EQ-5D-5L AUC = 0.67