French translation and validation of the Keele STarT MSK Tool

Charlotte Beaudart, PhD[1], Lorédana Criscenzo, MPT[2], Christophe Demoulin, PhD[2,4], Stephen Bornheim, PhD[1,2], Julien Van Beveren , PhD[5] and Jean-François Kaux, MD, PhD[1,2,3]

[1]Department of Physical Medicine and Rehabilitation, Liege University Hospital Center, University of Liège, Liège, Belgium, [2]Department of Rehabilitation and Sports Sciences, University of Liège, Liège, Belgium, [3]Department of Physical Medicine and Sports Traumatology, SportS(2), FIFA Medical Centre of Excellence, University and University Hospital of Liège, Liège, Belgium, [4]UCLouvain, Faculty of Motor Sciences at Université Catholique de Louvain-La-Neuve, Louvain, Belgium, [5]Haute Ecole de la ville de Liège (HEL), Liège, Belgium

ABSTRACT

Background: The Keele STarT MSK Tool is a 10-item questionnaire developed to classify patients suffering from one of the five most common types of musculoskeletal pain into three sub-groups according to their risk of chronic pain (i.e. low risk, medium risk and high risk). **Objectives**: The objective of the present study was to translate the Keele STarT MSK Tool into French and to evaluate its main psychometric properties. **Methods**: The translation and intercultural adaptation of the questionnaire were carried out using a 6-step process. The following psychometric properties were investigated: floor and ceiling effects, construct validity, internal consistency and test-retest reliability including Standard Error of Measurement and Smallest Detectable Change. **Results**: 101 patients suffering from musculoskeletal pain participated in the study. No floor nor ceiling effects were observed. A Cronbach's alpha of 0.65 was found, revealing moderate internal consistency. All items were demonstrated to be significantly correlated with the total score (range of correlations: r=0.2 for item 7 to r=0.78 for item 1). A significant correlation of r=0.78 between the French Keele STarT MSK Tool and the ÖMPSQ-short was found. Nevertheless, a poor agreement between tools was found, highlighted by a Kappa value of 0.57. Test-retest reliability was excellent (Intraclass Correlation Coefficient 0.97). The Standard Error of Measurement and Smallest Detectable Change of ±1.17 were 0.42 and ±1.17, respectively. **Conclusion**: A validated French version of the Keele STarT MSK Tool is now available and can be used by health practitioners to stratify patients as being low, medium or high risk for persistent musculoskeletal pain.

KEYWORDS: Chronic pain, Musculoskeletal pain, Translation, Validation Study.

Introduction

Musculoskeletal (MSK) conditions represent a considerable problem worldwide [1, 2]. MSK pain is prevalent and can place a heavy burden for those affected [3]. Moreover, consultations rates for MSK pain account for around a fifth of all consultations in primary care [4]. MSK conditions therefore have a major impact, not only for the individual but also on the health care system and society.

Patients with MSK pain in different body regions share common prognostic factors, as highlighted in a recent systematic review [5]. Therefore, prognostic stratification with recommended matched treatments for each subgroup has been suggested as clinically relevant and cost-effective. The Start Back Screening Tool (SBST), developed and validated in 2008 by Hill et al. [6, 7], is a brief questionnaire designed to screen primary care patients with low back pain for prognostic indicators that are relevant to initial decision making. This tool became rapidly very popular. Indeed, with a simple tool, patients are matched to treatment packages appropriate for them, which significantly decrease disability from low back pain, reduce time off work and save money by making better use of health resources.

Recently, a modified, generic version of the SBST, has been developed by Campbell et al. [8] for other musculoskeletal conditions. This new tool, the Keele STarT MSK Tool, allows for the stratification of primary

Corresponding author:
Dr. Charlotte Beaudart, CHU Liège, Service de médecine de l'appareil locomoteur, Bât. CNRF Centre Neurologique et de Réadaptation Fonctionnelle, Rue Champ des Alouettes 30, 4557 Tinlot, BELGIUM e-mail: c.beaudart@uliege.be

care patients suffering from one of the five most common types of musculoskeletal pain (back, neck, shoulder and multi-site pain) into three sub-groups at risk of chronic pain (low-, medium-and high risk). This tool has been designed to assist clinicians in their clinical decision making, by giving information on the likely risk of a poor outcome at the time a patient consults [9, 10]. The Keele STarT MSK Tool is comprised of 10 items assessing individual's function and disability, pain and coping, comorbidity and the impact of pain. The first item is composed of a 0-10 numeric rating scale (NRS) for pain and the remaining nine items consist of yes/no questions. The total score of the Keele STarT MSK Tool ranges therefore from 0 to 12 points. Participants with a score between 0 and 4 points, 5 and 8 points or 9 and 12 points are respectively considered at low, moderate or high risk of chronic pain [11].

The Keele STarT MSK Tool has shown good predictive and discriminative ability in development and validation samples [11, 12]. The Keele STarT MSK Tool was initially developed in English and, to our knowledge, has only been translated into Dutch [13]. Despite French being one of the most widespread languages worldwide and the fifth most spoken language, no French version of the Keele STarT MSK Tool is available. The objective of this study was therefore to translate the Keele STarT MSK Tool into French and to investigate its psychometric properties so as to confirm its validity and reliability as an instrument to classify risk of chronic pain for primary care patients suffering from MSK pain.

Methods

The study was developed using two main steps. First, the Keele STarT MSK Tool was translated in French and culturally adapted. Then, the questionnaire was submitted to a sample of individuals suffering from musculoskeletal pain for measuring psychometric properties of the French Keele STarT MSK Tool, respectively its internal consistency, test-retest reliability, construct validity and floor/ceiling effects. The study of the psychometric properties was performed considering the principles of the Consensus-based Standards for the Selection of Health Status Measurement Instruments (COSMIN) recommendations [14].The full protocol of the study is available under request to the corresponding author. The study protocol was approved by the Ethical Committee of the University of Liège.

French Translation and Adaptation Process

Permission was obtained from the developers of the Keele STarT MSK Tool to translate the questionnaire. The translation of the Keele STarT MSK Tool from English to French was performed according to guidelines of Guillemin et al.[15] and Beaton et al.[16]. This translation consisted of 6 phases: 1) during the first phase, the "forward-translation", two bilingual translators, native French-speakers, independently translated the tool into French; 2) during the second phase, a consensual version was produced by the two native French speakers; 3) during the third phase, the "back translation", two bilingual translators, native English-speakers, independently translated the common version back into English to check its accuracy; 4) during the fourth phase, an expert committee, composed of the four translators and a moderator, reviewed all of the documents, evaluated the conceptual differences between the original questionnaire and the back-translated version and produced a first version of the translated tool; 5) during the fifth phase the translated Keele STarT MSK Tool was subsequently presented to a linguist, to make sure that the translation was appropriate; 6) finally, during the last phase, the tool was submitted to 30 participants for a pre-test.

Study population

Candidates were eligible if they were adults and if they had recently consulted a doctor (either their general practitioner or their specialist)

for knee, cervical, scapular or diffuse musculoskeletal pain. Individuals suffering from cancer, inflammatory arthritis, spondylarthritis, neurological disease or individuals with recent trauma, cognitive impairment, dementia, terminal illness or suffering from pain related to pregnancy were excluded. Participants were recruited at the University Hospital of Liège (Belgium), at the Mont-Falise Hospital (Belgium) and also by means of social media and flyers distributed in gyms, physical therapy practices and companies/industries. Individual consent to participate was obtained from participants.

At the beginning of the study, socio-demographics characteristics as well as the site of pain of participants were recorded. Once included, participants were asked to complete the French Keele STarT MSK Tool twice, with a 7-day interval between. They were also invited to complete two additional questionnaires. First, they completed the French validated version of the Örebro Musculoskeletal Pain Screening Questionnaire Short-Form (ÖMPSQ-short) [17, 18], which is a 10-item questionnaire that identifies workers at a higher risk of work disability. The total score obtained varies from 1 to 100 points, with a score higher than 50 reflecting a higher estimated risk for future work disability [19]. Second, they completed the French version of the Short-Form 12 (SF-12) questionnaire [20], which is a generic quality of life questionnaire. The SF-12 contains a mental component (MCS) and a physical component (PCS) both are scored out of 100 points.

Participants were categorized as being at low, moderate or high risk of chronic MSK pain according to the results of the French Keele STarT MSK Tool. Differences of sociodemographic and clinical characteristics between these three categories were measured using an ANOVA for continuous variables that followed a normal distribution, a Kruskal-Wallis test for continuous variables that did not follow a normal distribution and with a Chi squared test or Fisher exact test for qualitative variables.

Psychometric evaluations

All analyses were carried out with IBM SPSS for Windows, version 25 (Armonk, NY: IBM Corp.). Normality of distribution of continuous variables was first established on the basis of the distance between mean and median, the histogram, the quantile-quantile plot and the Shapiro-Wilk test. Variables that displayed normal distribution were reported as mean and standard deviation, and non-normal variables as median and interquartile range (P25–P75). Qualitative and binary variables were expressed as absolute and relative frequencies. Moreover, the choice of psychometric statistics used in the analyses was made according to the distribution of variables.

Internal consistency: The estimation of homogeneity across items of the French Keele STarT MSK Tool, or internal consistency, was analysed by calculating the Cronbach's alpha of the total score. The value of alpha varies between 0 and 1, with a good level of internal consistency established when the alpha value ranges from 0.70 to 0.95 [21].The questionnaire's internal consistency was evaluated as a whole and by deleting a single item one by one. Moreover, to further test internal consistency, we also measured correlations between the total score and each individual item. Significant correlations higher than 0.6 were considered as strong correlations [22].

Construct validity: The construct validity of the French Keele STarT MSK Tool was analysed by correlation with the ÖMPSQ-short. Pearson's correlation coefficient was used for comparison with a high correlation expected between both instruments. Significant correlations higher than 0.6 were considered as strong corelations [22]. Moreover, a Cohen Kappa Coefficient was measured to reflect agreement between the Keele STarT MSK Tool and the ÖMPSQ-short for the categorization of patients at high risk of chronic pain. A Cohen Kappa coefficient higher than 0.6 was considered as acceptable [23].

Floor/ceiling effects: Floor and ceiling effects were considered to be present if more than 15% of the population obtained a maximum score (ceiling effect) or a minimum score (floor effect).

Test-retest reliability: Test-retest reliability shows the extent to which the questionnaire produces the same scores for repeated measurements in participants whose health has not changed [24]. For the one-week test-retest reliability, the Intraclass Correlation Coefficient (ICC – two-way mixed, absolute agreement) was calculated for the total score of the questionnaire. ICC values higher than 0.7 were considered as acceptable [25]. Cohen's Kappa coefficient was calculated for individual binary items, with values higher than 0.6 considered as acceptable [23]. The standard error of measurement (SEM) and the smallest detectable change (SDC) of the questionnaire were also calculated using the following formulas [21]: SEM= SD * square root (1-ICC) SDC= 1.96 * SEM * square root (2) The SEM provides a range around the observed value in which the theoretical "true" value lies. The SDC indicates the amount of change that needs to be measured to be sure that the change measured is real, and not potentially a product of measurement error [24].

Results

French Translation and Adaptation Process
The 10 questions of the Keele STarT MSK Tool were translated without any major difficulties. All differences between translators, both in the phases of translation and back-translation, were resolved by consensus. A prefinal version of the translated questionnaire was pre-tested with 30 participants. The comprehensibility of the questionnaire was excellent, no further change was required during this pilot phase. The final French version of the tool has been validated by the expert committee and is available in **Appendix 1**.

Psychometric evaluation
Study population: 101 participants took part in this study and completed both the French Keele STarT MSK Tool and the ÖMPSQ-short tools. The sample consisted of 63 women and 38 men with a median age of 25.0 years (interquartile range 21.5-41.0 years). The majority of the sample was composed of active workers (44.6%) or students (34.7%). Respectively 29, 54 and 18 participants were categorized as low, moderate and high risk of chronic pain according to the Keele STarT MSK Tool. Back pain was the most frequently reported symptom by the participants (26.7% of the sample), followed by knee pain (24.8% of the sample). Based on the first item of the Keele STarT MSK Tool, pain was significantly lower for individuals categorized as low risk of chronic pain (p<0.001). Moreover, a better quality of life, both for the mental (p=0.01) and physical component scales (p<0.001) of the SF-12 was observed for individuals at low risk of chronic pain (**Table 1**).

Internal consistency: A Cronbach's alpha of 0.65 was found revealing moderate internal consistency. When one item was removed from the analysis, the lowest Cronbach's alpha obtained was 0.55 (when removing item 1). Removing other items did not affect the internal consistency of the French Keele STarT MSK Tool (**Table 2**). All items also demonstrated significant correlations with the total score (highest correlation for item 1 with r=0.78, p<0.001, lowest correlation for item 7 with r=0.2, p=0.04).

Construct validity: The French Keele STarT MSK Tool had good construct validity highlighted by a strong and significant correlation with the ÖMPSQ-short (r=0.78, p<0.001). The ÖMPSQ-short score was significantly lower for participants at low risk of chronic pain, as diagnosed with the Keele STarT MSK Tool (28 ± 10.8 points vs 42.1 ± 11.5 points vs 65 ± 11.9 points for individuals at low, moderate and high risk of chronic pain, p<0.001, respectively) (**Table 1**). Using the Keele STarT MSK Tool, 18 (18%) of the participants were categorized as "high risk", using the

ÖMPSQ-short, 30 (30%) were categorized as "high risk". Agreement between both tools was not optimal, however, with a kappa coefficient of 0.57, (95% CI 0.39 ; 0.75, p<0.001).

Floor and Ceiling effects: No floor, nor ceiling effect was observed for the French Keele STarT MSK Tool.

Test-retest reliability Among the 101 participants, 51 participants did not report any modification of pain during the 7-day interval and could therefore be invited to complete the French Keele STarT MSK Tool a second time. Test-retest reliability for those 51 participants for the total score was excellent with an ICC of 0.97 (95% CI 0.95-0.98). Kappa Coefficient values for individual items were excellent and ranged from 0.79-0.95 (**Table 3**). Using this sample of 51 participants included in reliability analyses, a SEM of 0.42 and a SDC of 1.17 were obtained.

Discussion

Evidence shows that chronic MSK pain represents a considerable burden for the individual and society. It is essential to have tools to categorize individuals into subgroups to offer matched and appropriate treatments with regards to the clinical conditions of patients and avoid pain becoming chronific. Our study has established a French version of the Keele STarT MSK Tool with acceptable internal consistency, excellent test-retest reliability and absence of floor/ceiling effects. Although we found a good correlation between the total scores of the French Keele STarT MSK Tool and the French ÖMPSQ-short tool, there was poor agreement between both tools in identifying patients at high risk of chronic pain, highlighted by a Kappa coefficient of 0.57.

To provide equivalence between the English and the new French version of the Keele STarT MSK Tool, a rigorous translation and cross-cultural adaptation process was followed. Input from the 30 participants of the pre-test and from a linguistic expert during the expert committee review confirmed that the French Keele STarT MSK Tool has the same content as the original English Keele STarT MSK Tool, while also being comprehensible to its target audience.

The sample of participants included in the validity study consisted of 101 individuals suffering from one of the five most common types of musculoskeletal pain. In accordance with the literature, highlighting a higher proportion of women in medical consultations for MSK pain [26], we also included a higher proportion of women in the study (62.4%). In our sample, the Keele STarT MSK Tool classified 18% of the population as high risk for chronic pain, which is in between the prevalence found in the study by Dunn et al. [11] who validated the original Keele STarT MSK Tool (i.e. 33% of the population was categorized as "high risk") and the prevalence found in the study of van den Broek13 et al. who validated the Dutch translation of the Keele STarT MSK Tool (i.e. 2.8% of the population was categorized "high risk"). In the Dutch validation study13, authors raised some hypotheses to explain the low prevalence found. Among the hypotheses, they explained that they included only patients that consulted their physiotherapist and not their general practitioners, in contrast with the Dunn et al. [11] study. In our study, we included participants consulting their general practitioners or addressed to a physiotherapist by a specialist physician which might explain that our prevalence was higher than in Van den Broek et al. 13 study.

Psychometric analyses revealed that the French Keele STarT MSK Tool has a moderate but acceptable internal consistency (Cronbach's alpha of 0.65). With this value, the internal consistency of the Keele STarT MSK Tool could not be considerate as excellent. However, all items appear to be significantly and positively correlated with the total score of the questionnaire, revealing that, even if homogeneity of the tool is not perfect, it is nevertheless acceptable for its purpose. Only item 7 demonstrated a relatively low correlation with the total score. Neither the validation of the

Table 1 Socio-demographic and clinical characteristics of the population

	Total sample n=101	Low KSMT (0-4 points) n=29	Moderate KSMT (5-8 points) n=54	High KSMT (9-12 points) n=18	p-values*
Sex					
Women	63 (62.4)	15 (51.7)	37 (68.5)	11 (61.1)	0.32
Age (years)	25.0 (21.5-41.0)	24.5 (21.0-27.1)	27.0 (22.7-41.7)	24.5 (20.7-41.2)	0.29
Professional status					
Active	45 (44.6)	9 (31.0)	28 (51.9)	8 (44.4)	0.48
Student	35 (34.7)	12 (41.4)	16 (29.6)	7 (38.9)	
Incapacity	8 (7.9)	2 (6.8)	4 (7.4)	2 (11.1)	
Retired	6 (5.9)	2 (6.9)	3 (5.6)	1 (5.6)	
Unemployed	7 (6.9)	4 (13.8)	3 (5.6)	0 (0.0)	
BMI (kg/m²)	24.1 (20.9-26.4)	24.5 (21.0-27.1)	24.0 (20.7-26.2)	23.9 (20.6-27.1)	0.95
Site of pain					
Neck	15 (14.9)	4 (13.8)	8 (14.8)	3 (16.7)	0.14
Back	27 (26.7)	6 (20.7)	16 (29.6)	5 (27.8)	
Shoulder	18 (17.8)	5 (17.2)	12 (22.2)	1 (5.6)	
Knee	25 (24.8)	12 (41.4)	10 (18.5)	3 (16.7)	
Multi-site	16 (15.8)	2 (6.9)	8 (14.8)	6 (33.3)	
Pain (NRS)	5.64 ± 2.05	3.59 ± 1.35	5.96 ± 1.47	8.00 ± 1.24	<0.001
SF-12					
PCS (/100)	50.0 ± 9.52	56.5 ± 7.52	50.0 ± 7.55	39.4 ± 8.44	<0.001
MCS (/100)	50.0 ± 10.3	51.1 ± 10.9	51.3 ± 9.94	44.2 ± 9.03	0.01
ÖMPSQ-short	42.1 ± 16.7	28.0 ± 10.8	42.1 ± 11.5	65.0 ± 11.9	<0.001
KSMT (0/12)	5.74 ± 2.56	2.6 ± 1.18	6.2 ± 0.99	9.5 ± 0.71	<0.001

KSMT: Keele STarT MSK tool ; BMI: Body Mass Index ; ÖMPSQ-short: Örebro Musculoskeletal Pain Screening Questionnaire Short-Form; NRS: Numeric Rating Scale *p-values for continuous variables obtained from an ANOVA test for continuous variables with a normal distribution (ÖMPSQ-short, Pain(NRS), SF-12 and KSMT) and from a Kruskal-Wallis test for continuous variables without a normal distribution (age and BMI) ; p-values for qualitative variables obtained from a Chi² test or a Fisher exact test.

Table 2 Values of internal consistency

	Cronbach's alpha when item removed	Correlation with the total score of the KSMT	P-value for correlation
Item 1	0.55	0.79	<0.001
Item 2	0.61	0.53	<0.001
Item 3	0.61	0.54	<0.001
Item 4	0.66	0.28	0.005
Item 5	0.63	0.44	<0.001
Item 6	0.62	0.47	<0.001
Item 7	0.66	0.20	0.04
Item 8	0.58	0.64	<0.001
Item 9	0.63	0.44	<0.001
Item 10	0.63	0.44	<0.001

KSMT: Keele STarT MSK tool

Table 3 Kappa Coefficient values for each individual items

Item	Kappa	95%CI
Item 1	0.81	0.70 ; 0.94
Item 2	0.89	0.74 ; 1.04
Item 3	0.95	0.85 ; 1.05
Item 4	0.89	0.74 ; 1.04
Item 5	0.79	0.56 ;1.01
Item 6	0.95	0.84 ; 1.05
Item 7	0.93	0.81 ; 1.06
Item 8	0.91	0.79 ; 1.03
Item 9	0.91	0.78 ; 1.03
Item 10	0.95	0.86 ; 1.04

original tool [11] nor the validation of the Dutch translation of the Keele STarT MSK Tool13, provided a measure of internal consistency of the tool, which prevents us from comparison. During the French translation and validation of the SBST, the authors found a Cronbach's alpha of 0.74 for the psychological subscale of the tool, which is close, but nevertheless better than the internal reliability of the Keele STarT MSK Tool. However, the Keele STarT MSK Tool does not have a psychological subscale, which does not allow for optimal comparison between the tools.

Construct validity of the French Keele STarT MSK Tool has been demonstrated by a strong, positive and significant correlation with the ÖMPSQ-short questionnaire [27]. Nevertheless, when measuring agreement between tools to classify individuals as "high risk" of chronic pain, we found poor concordance, with a Kappa coefficient of 0.57, which is below the acceptable threshold of 0.6. This apparently opposite result simply reflects that both tools seem to evolve in the same direction and are therefore correlated but, unfortunately, they do not classify the same individuals as high risk.

We also demonstrated excellent test-retest reliability over a one-week period, both for individual items and for the total score of the French Keele STarT MSK Tool. For this analysis, only half of the participants completed the French Keele STarT MSK Tool twice. Even if the sample was restricted compared to the full sample of participants, the sample size is acceptable from a methodological point of view16. The validation of the original Keele StarT MSK tool did not provide a reliability analysis. However, in the Dutch translation and validation study, authors found a quadratic-weighted kappa coefficient for the total score of 0.71. Because of the continuous format of the total score of the questionnaire, we decided to measure an ICC for the total score and measure Kappa Coefficients for all items separately. We found high values of reliability with an ICC of 0.97 for the total score and Kappa values higher than 0.79 for individual items, which confirmed high reliability of the questionnaire. In our study, we were also able to calculate the standard error of measurement and minimal detectable change of the questionnaire, which was not computed in prior studies. The MDC for the French Keele STarT MSK Tool is 1.17 points (on a scale from 0 to 12 points), which means that a individuals would have to change by at least this amount before we can be sure that he/she has actually improved or deteriorated.

Finally, no floor nor ceiling effects were identified. This psychometric property is important for the capacity of the tool to detect changes. When floor/ceiling effects are identified, tools are less sensitive to change and less able to discriminate between patients.

Limitations

Because of the cross-sectional design of our study, we were unable measure the responsiveness, the minimal clinical important difference nor the predictive validity of the French Keele STarT MSK Tool. During their Dutch translation process, van de Broek et al.13 measured the ability of the Dutch Keele STarT MSK Tool to predict persisting disability at 3 months. They found relative risk ratios for persisting disability of 2.2 for the medium risk group and 7.3 for the high-risk group. However, even if they were the first ones to provide longitudinal psychometric properties of this questionnaire, their sample size for high-risk patients only included 4 patients, which limits generalisability of findings in this group. Still because of the cross-sectional design of the study, we measured the validity of the French Keele STarT MSK tool only through construct validity and not using criterion validity, which is also questionable. Indeed, it would have been very informative to obtain longitudinal data on the incidence of chronic pain in our population and the possible association between this incidence and the results obtained from the Keele STarT MSK tool. Another possible limitation of our study is that almost 80% of our sample was comprised of active workers and students. A selection bias may have occurred because we mainly recruited via social networks and flyers in companies/industries. Therefore, our sample is represented by a high proportion of active workers, and a lower proportion of unemployed or retired participants, or people with work incapacities.

Perspective

Results of our study encourages clinicians to use the French Keele STarT MSK tool to help identify French-speaking patients at risk of chronicity and to identify the presence of psychosocial factors to be taken into account in the clinical management of the patient. Because the responsiveness, the minimal clinical important difference, and the criterion validity of the questionnaire have never been measured, further longitudinal studies using the Keele STarT MSK tool should be encouraged.

Conclusion

The French version of the Keele STarT MSK Tool is now available and shows good validity and reliability values. This tool can be used with confidence to obtain a classification of individuals suffering from MSK conditions at low, moderate or high risk of chronic pain. Before this study, the Keele STarT MSK Tool had been only validated in one unique

population study. With this study, we validated it in a second cohort from a different country.

Disclosure of interest

The authors report no conflict of interest

Ethics

The study was approved by the Ethical Committee of the University of Liège

Data availability

All data are available under request to the corresponding author

Authors' contribution

J-F. Kaux, and C. Demoulin designed the study. L. Criscenzo recruited participants and collected the data. S. Bornheim and J. van Beveren were translators and participated in the interpretation of results. L. Criscenzo and C. Beaudart run the analyses. C. Beaudart, J-F. Kaux and C. Demoulin interpreted the data. C. Beaudart drafted the manuscript. All authors reviewed the manuscript and approved it.

Acknowledgments

Fundings

References

[1] Christopher JL Murray, Theo Vos, Rafael Lozano, Mohsen Naghavi, Abraham D Flaxman, Catherine Michaud, Majid Ezzati, Kenji Shibuya, Joshua A Salomon, Safa Abdalla, et al. Disability-adjusted life years (dalys) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the global burden of disease study 2010. *The lancet*, 380(9859):2197–2223, 2012. doi: 10.1016/S0140-6736(12)61689-4.

[2] Fiona M Blyth, Andrew M Briggs, Carmen Huckel Schneider, Damian G Hoy, and Lyn M March. The global burden of musculoskeletal pain—where to from here? *American journal of public health*, 109(1):35–40, 2019. doi: 10.2105/AJPH.2018.304747.

[3] Hanneke AH Wijnhoven, Henrika CW De Vet, and H Susan J Picavet. Prevalence of musculoskeletal disorders is systematically higher in women than in men. *The Clinical journal of pain*, 22(8):717–724, 2006. doi: 10.1097/01.ajp.0000210912.95664.53.

[4] Kelvin P Jordan, Umesh T Kadam, Richard Hayward, Mark Porcheret, Catherine Young, and Peter Croft. Annual consultation prevalence of regional musculoskeletal problems in primary care: an observational study. *BMC musculoskeletal disorders*, 11(1):1–10, 2010. doi: 10.1186/1471-2474-11-144.

[5] Majid Artus, Paul Campbell, Christian D Mallen, Kate M Dunn, and Danielle AW van der Windt. Generic prognostic factors for musculoskeletal pain in primary care: a systematic review. *BMJ open*, 7(1):e012901, 2017. doi: 10.1136/bmjopen-2016-012901.

[6] Jonathan C Hill, David GT Whitehurst, Martyn Lewis, Stirling Bryan, Kate M Dunn, Nadine E Foster, Kika Konstantinou, Chris J Main, Elizabeth Mason, Simon Somerville, et al. Comparison of stratified primary care management for low back pain with current best practice (start back): a randomised controlled trial. *The Lancet*, 378(9802):1560–1571, 2011. doi: 10.1016/S0140-6736(11)60937-9.

[7] Jonathan C Hill, Kate M Dunn, and Martyn Lewis. A primary care back pain screening tool: identifying patient subgroups for initial treatment. *Arthritis rheum*, 59(5):632–641, 2008. doi: 10.1002/art.23563.

[8] Paul Campbell, Jonathan C Hill, Joanne Protheroe, Ebenezer K Afolabi, Martyn Lewis, Ruth Beardmore, Elaine M Hay, Christian D Mallen, Bernadette Bartlam, Benjamin Saunders, 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. *Journal of pain research*, 9:807, 2016. doi: 10.2147/JPR.S116614.

[9] Benjamin Saunders, Jonathan C Hill, Nadine E Foster, Vince Cooper, Joanne Protheroe, Adrian Chudyk, Carolyn Chew-Graham, and Bernadette Bartlam. Stratified primary care versus non-stratified care for musculoskeletal pain: qualitative findings from the start msk feasibility and pilot cluster randomized controlled trial. *BMC family practice*, 21(1):1–12, 2020. doi: 10.1186/s12875-020-1098-1.

[10] JC Hill, S Garvin, Y Chen, V Cooper, S Wathall, B Saunders, M Lewis, J Protheroe, A Chudyk, KM Dunn, 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 family practice*, 21(1):1–18, 2020.

[11] Kate M Dunn, Paul Campbell, Martyn Lewis, Jonathan C Hill, Danielle A van der Windt, Ebenezer Afolabi, Joanne Protheroe, Simon Wathall, Sue Jowett, Raymond Oppong, et al. Refinement and validation of a tool for stratifying patients with musculoskeletal pain. *European Journal of Pain*, 2021. doi: 10.1002/ejp.1821.

[12] JC Hill, EK Afolabi, M Lewis, KM Dunn, E Roddy, DA Van Der Windt, and NE Foster. Does a modified start back tool predict outcome with a broader group of musculoskeletal patients than back pain? a secondary analysis of cohort data. *BMJ open*, 6(10):e012445, 2016. doi: 10.1136/bmjopen-2016-012445.

[13] Anke G van den Broek, Corelien JJ Kloek, Martijn F Pisters, and Cindy Veenhof. Validity and reliability of the dutch start msk tool in patients with musculoskeletal pain in primary care physiotherapy. *PloS one*, 16(3):e0248616, 2021. doi: 10.1371/journal.pone.0248616.

[14] Lidwine B Mokkink, Cecilia AC Prinsen, Lex M Bouter, Henrica CW de Vet, and Caroline B Terwee. The consensus-based standards for the selection of health measurement instruments (cosmin) and how to select an outcome measurement instrument. *Brazilian journal of physical therapy*, 20:105–113, 2016. doi: 10.1590/bjpt-rbf.2014.0143.

[15] Francis Guillemin, Claire Bombardier, and Dorcas Beaton. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *Journal of clinical epidemiology*, 46(12):1417–1432, 1993. doi: 10.1016/0895-4356(93)90142-N.

[16] Dorcas E Beaton, Claire Bombardier, Francis Guillemin, and Marcos Bosi Ferraz. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*, 25(24):3186–3191, 2000. URL http://www.ncbi.nlm.nih.gov/pubmed/11124735.

[17] Charles Philip Gabel, Markus Melloh, Brendan Burkett, Jason Osborne, and Michael Yelland. The örebro musculoskeletal screening questionnaire: validation of a modified primary care musculoskeletal screening tool in an acute work injured population. *Manual therapy*, 17(6):554–565, 2012. doi: 10.1016/j.math.2012.05.014.

[18] R Hilfiker, IA Knutti, B Raval-Roland, G Rivier, Geert Crombez, and Emmanuelle Opsommer. Validity and responsiveness of the french version of the örebro musculoskeletal pain screening questionnaire in chronic low back pain. *European Spine Journal*, 25(9): 2741–2749, 2016. doi: 10.1007/s00586-016-4635-9.

[19] Steven J Linton, Michael Nicholas, and Shane MacDonald. Development of a short form of the örebro musculoskeletal pain screening questionnaire. *Spine*, 36(22):1891–1895, 2011. doi: 10.1097/BRS.0b013e3181f8f775.

[20] Crispin Jenkinson and Richard Layte. Development and testing of the uk sf-12. *Journal of health services research & policy*, 2(1):14–18, 1997. doi: 10.1177/135581969700200105.

[21] Henrica CW De Vet, Caroline B Terwee, Lidwine B Mokkink, and Dirk L Knol. *Measurement in medicine: a practical guide*. Cambridge university press, 2011. doi: 10.1017/CBO9780511996214.

[22] Richard A Deyo, Paula Diehr, and Donald L Patrick. Reproducibility and responsiveness of health status measures statistics and strategies for evaluation. *Controlled clinical trials*, 12(4):S142–S158, 1991. doi: 10.1016/S0197-2456(05)80019-4.

[23] Mary L McHugh. Interrater reliability: the kappa statistic. *Biochemia medica*, 22(3):276–282, 2012. doi: 10.11613/bm.2012.031.

[24] Lidwine B Mokkink, Caroline B Terwee, Donald L Patrick, Jordi Alonso, Paul W Stratford, Dirk L Knol, Lex M Bouter, and Henrica CW De Vet. The cosmin checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international delphi study. *Quality of life research*, 19(4):539–549, 2010. doi: 10.1007/s11136-010-9606-8.

[25] Caroline B Terwee, Sandra DM Bot, Michael R de Boer, Daniëlle AWM van der Windt, Dirk L Knol, Joost Dekker, Lex M Bouter, and Henrica CW de Vet. Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of clinical epidemiology*, 60(1):34–42, 2007. doi: 10.1016/j.jclinepi.2006.03.012.

[26] Theo Vos, Christine Allen, Megha Arora, Ryan M Barber, Zulfiqar A Bhutta, Alexandria Brown, Austin Carter, Daniel C Casey, Fiona J Charlson, Alan Z Chen, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the global burden of disease study 2015. *The lancet*, 388(10053):1545–1602, 2016. doi: 10.1016/S0140-6736(16)31678-6.

[27] Anna Sofia Simula, Olli Ruokolainen, Petteri Oura, Mikko Lausmaa, Riikka Holopainen, Maija Paukkunen, Juha Auvinen, Steven J Linton, Jonathan C Hill, and Jaro Karppinen. Association of start back tool and the short form of the örebro musculoskeletal pain screening questionnaire with multidimensional risk factors. *Scientific reports*, 10(1):1–11, 2020. doi: 10.1038/s41598-019-57105-3.