Construct validity and test-retest reliability of the Dutch STarT MSK tool in patients with musculoskeletal pain in primary care physiotherapy

Masterthesis
Physiotherapy Science
Program in Clinical Health Sciences
Utrecht University

Name student: A.G. (Anke) van den Broek
Student number: 5665140
Date: 28 June 2018
Internship supervisor(s): Dr. C.J.J. Kloek, Prof. dr. C. Veenhof
Internship institute: Department of Rehabilitation, Physical Therapy Science and Sports, Rudolf Magnus Institute of Neurosciences, University Medical Centre Utrecht, Utrecht University, Utrecht, The Netherlands
Lecturer/supervisor Utrecht University: Dr. J. van der Net
"ONDERGETEKENDE

Anke Gertruda van den Broek,

bevestigt hierbij dat de onderhavige verhandeling mag worden geraadpleegd en vrij mag worden gefotokopieerd. Bij het citeren moet steeds de titel en de auteur van de verhandeling worden vermeld."
Examiner:
Dr. M.F. Pisters

Assessors:
Prof. dr. C. Veenhof
Dr. M.F. Pisters

Masterthesis, Physical Therapy Sciences, Program in Clinical Health Sciences, Utrecht University, Utrecht, 2018
ABSTRACT

Background The STarT MSK tool is a risk stratification tool, recently developed and validated in the United Kingdom to allocate primary care patients with musculoskeletal pain into three prognostic subgroups. A first step toward identifying whether the STarT MSK tool could be useful for Dutch clinical practice as well, is to evaluate its measurement properties.

Aim To evaluate the construct validity and test-retest reliability of the Dutch STarT MSK tool in patients with musculoskeletal pain in primary care physiotherapy.

Methods A cross-sectional observational design was used to evaluate construct validity and a ‘5-day’ test-retest design to measure test-retest reliability. Physiotherapists included patients with musculoskeletal pain, aged 18 years or older. Patients completed a questionnaire at baseline and after 5 days. Construct validity was assessed by comparing scores of the STarT MSK items with reference questionnaires. Pearson’s correlation coefficients were calculated to test predefined hypotheses. Test-retest reliability was evaluated by calculating quadratic-weighted kappa coefficients for the overall tool scores and prognostic subgroups.

Results In total, 142 patients were included for psychometric analysis. At baseline, 74 patients (52.1%) were categorised as low risk, 64 (45.1%) as medium risk and 4 (2.8%) as high risk. For construct validity, nine of the eleven predefined hypotheses were confirmed. The correlations between item 2 (pain self-efficacy) and item 9 (kinesiophobia) with their respective reference questionnaires were lower than expected. The kappa coefficients for the overall tool scores and prognostic subgroups were 0.707 and 0.653, respectively. Test-retest reliability increased to 0.746 for the overall tool scores and decreased to 0.603 for prognostic subgroups, when agreement was calculated in the subset of 72 patients reporting stable musculoskeletal pain symptoms.

Conclusion The Dutch STarT MSK tool showed a good construct validity and substantial test-retest reliability in patients with musculoskeletal pain in primary care physiotherapy.

Implications of key findings This validation study demonstrates promising results regarding the measurement properties of the Dutch STarT MSK tool. However, further psychometric evaluation concerning the predictive validity and additional analyses across pain sites are required before the tool can be implemented in clinical practice.

Keywords Musculoskeletal pain, physiotherapy, classification, validity, reliability.
INTRODUCTION

Musculoskeletal conditions, such as low-back pain, neck pain, osteoarthritis and rheumatoid arthritis, are the most common cause of long-term pain and impaired physical function, and have large impact on health-related quality of life. In developed countries, musculoskeletal conditions represent the third greatest health burden of the population. In the Netherlands the point prevalence of musculoskeletal pain is found to be 53.9%, with low-back, shoulder, neck and knee being the most frequently affected pain sites. Musculoskeletal conditions have high impact on health care costs, being responsible for the second highest healthcare costs among all chronic diseases within the Netherlands. Furthermore, chronic musculoskeletal conditions are associated with high work-related costs due to work absenteeism and impaired work performance.

Musculoskeletal pain is predominantly managed in primary care, for example by the physiotherapist. Although current evidence shows positive effects of exercise therapy and psychosocial interventions on pain and function in patients with musculoskeletal pain, it needs to be considered that every patient is unique and some patients respond better to certain treatments than others. In order to improve effectiveness on clinical outcomes and cost-effectiveness in the treatment of patients with musculoskeletal pain, a stratified care approach is promising. Within stratified care, treatments are matched to patients based on prognostic information such as biomedical and psychosocial risk factors for poor prognosis. To identify modifiable risk factors for poor prognosis at an early stage and, subsequently, to stimulate that the appropriate stratified care will be applied to patients, a valid and reliable risk stratification tool is required.

The Keele STarT Back-Screening Tool (SBT) is an example of a valid and reliable risk stratification tool developed to allocate primary care patients with low-back pain into three prognostic subgroups (low, moderate and high risk of persisting back pain disability), and to apply the appropriate matched treatment. The original SBT was developed in the United Kingdom (UK) and has been translated into several languages, including Dutch. A stratified care approach (use of the SBT and matched treatments) has demonstrated superior clinical and cost outcomes compared to usual non-stratified primary care in patients with low-back pain within the UK. While the SBT focuses primarily on back pain, there is evidence that different regional musculoskeletal pain presentations share common underlying mechanisms and prognostic factors. These studies indicate that a comparable risk stratification tool could be useful for patients with a broader range of musculoskeletal pain presentations.

Recently, a modified, generic version of the SBT was developed for patients with the five most common musculoskeletal pain presentations (i.e., neck, back, shoulder, knee or multisite pain) within the UK. This so-called Keele STarT MSK tool showed a moderate to good predictive ability of the tool’s baseline score for identifying patients who developed persisting disability because of musculoskeletal pain, and subgroup cut-points were comparable across pain sites. Given the promising predictive performance of the STarT MSK tool, we believe that the tool could be useful for Dutch clinical practice as well. A first step
toward identifying whether the STarT MSK tool could be useful for Dutch clinical practice, is to translate the tool into Dutch and evaluate its measurement properties.

The STarT MSK tool is a risk stratification tool using ten independent prognostic indicators for persisting disability because of musculoskeletal pain. As the prediction of persisting disability is based on these prognostic indicators, it is important to assess whether the specific items of the STarT MSK tool indeed measure the intended underlying biomedical and psychosocial factors. Furthermore, an instrument with predictive purposes has to be reliable (i.e., classify clinically stable individuals into the same prognostic subgroup over time), because low reliability can influence the predictive ability of the STarT MSK tool. Therefore, the aim of this study is to evaluate the construct validity and test-retest reliability of the Dutch version of the STarT MSK tool in patients with musculoskeletal pain in primary care physiotherapy.

METHODS

Translation of the STarT MSK tool
Prior to the start of this study, we formally translated the original English version of the STarT MSK tool (Appendix 1) into Dutch using a forward-backward translation method according to the guidelines of Beaton et al. Two native Dutch speakers independently performed a forward translation. Based on a consensus meeting a single preliminary Dutch version was formed. This version was translated back into English independently by two native English speakers with no medical background. An expert committee consisting of two forward translators, one backward translator, two clinical health scientists and one physiotherapist then reviewed the original STarT MSK tool and each translated version, which resulted in a pre-final version of the Dutch STarT MSK tool. Finally, this pre-final version was tested in a pilot consisting of 20 Dutch-speaking patients who consulted a physiotherapist for musculoskeletal pain. After completing the questionnaire, patients were briefly interviewed about the interpretation and comprehension of each item and the chosen response. As no problems were reported, the pre-final version of the Dutch STarT MSK tool was considered final. The Dutch version of the STarT MSK tool is included in Appendix 2.

Design
A cross-sectional observational design was used to evaluate construct validity and a ‘5-day’ test-retest design to measure test-retest reliability. Patients were asked to complete a baseline and follow-up questionnaire, and received usual care from their physiotherapist. This study was not subject to the Medical Research Involving Human Subjects Act (WMO) and received a ‘non-WMO’ declaration from the Medical Research Ethics Committee of the University Medical Centre Utrecht, The Netherlands (registration number 18-082).
Participants

Physiotherapists
A total of 65 primary care physiotherapists within the authors’ network were invited to participate in this study. Only physiotherapists who are seeing at least 1 to 2 patients with musculoskeletal pain for a first consultation per week were eligible to participate in this study.

Patients
The inclusion period for patients was February 2018 to May 2018. Patients were eligible for inclusion when (1) they consulted a physiotherapist for musculoskeletal pain (i.e., neck, back, shoulder, knee or multisite pain), (2) it was the first consultation for the current episode of musculoskeletal pain, (3) they were aged 18 years or older, (4) they were able to read and write in Dutch and (5) they had an email address. Patients were excluded when (1) during the first consultation red flags were found indicating a possible specific underlying pathology (e.g., fracture, infection, tumor, cauda equina) responsible for the musculoskeletal pain, (2) they consulted a physiotherapist for pre- or post-operative rehabilitation related to the musculoskeletal pain presentation, (3) they were diagnosed with inflammatory arthritis, spondyloarthropathy or polymyalgia rheumatica, (4) they experienced pregnancy-related pain problems or (5) they had a cognitive impairment.

Study procedure
The participating physiotherapists received information about the study procedure during a one-hour in company instruction. Physiotherapists informed eligible patients about the study and screened them on in- and exclusion criteria. Patients who were willing to participate received an information letter from the physiotherapist. After patients had given permission to be contacted, the researcher (AB) received their information from the physiotherapist. Patients were then e-mailed by the researcher (AB) and received a link to the informed consent form as part of the baseline questionnaire. The link to the baseline questionnaire was sent within 24 hours after the first consultation with the physiotherapist. Patients could only start filling-in the questionnaire if informed consent was provided. When necessary, a reminder was sent to the patient within a few days. The link to the follow-up questionnaire was sent 5 days after filling-in the baseline questionnaire.

Measurements
Baseline
At baseline (T0) patients received an online questionnaire, consisting of the following measurements:

- General patient characteristics including age, gender, educational level, pain site, pain duration and presence of comorbidity.
- The prediction of persisting musculoskeletal-related disability was assessed with the
Dutch STarT MSK tool, consisting of 10 independent items that cover biomedical and psychosocial prognostic factors (Figure 1). Subgroup cut-points are 0-4 for low risk, 5-8 for medium risk and 9-12 for high risk, based on an overall score ranging from 0-12 (P. Campbell, Research Associate Keele University, personal communication through e-mail, 15 March 2018).

- The average pain in the past week was measured with the 11-point Numeric Pain Rating Scale (NPRS),\textsuperscript{25} ranging from 0 (no pain) to 10 (worst possible pain). The NPRS is valid and reliable for use in clinical practice.\textsuperscript{26}
- Pain self-efficacy beliefs were assessed with the valid and reliable Pain Self-Efficacy Questionnaire (PSEQ),\textsuperscript{27} consisting of 10 items, each scored on a 7-point Likert scale (0 = not at all confident; 6 = completely confident), with a higher score reflecting stronger self-efficacy beliefs.
- Disability was measured with the valid and reliable physical functioning subscale of the 36-Item Short Form Health Survey (SF-36 PF),\textsuperscript{28} consisting of 10 statements with three answer options varying from ‘Yes, limited a lot’ to ‘No, not limited at all’. Each item is scored 1 to 3 points and the total score was transformed to a 100-point scale, with a higher score indicating better physical functioning.
- Timeline illness perception was assessed with the timeline question of the Brief Illness Perception Questionnaire (Brief IPQ),\textsuperscript{29} ranging from ‘a very short time’ to ‘forever’ on an 11-point scale.
- Depressive symptoms were measured with the valid and reliable depression subscale of the Hospital Anxiety and Depression Scale (HADS-D),\textsuperscript{30} consisting of 7 items, each scored on a 4-point Likert scale, with a higher score reflecting more depressive symptoms.
- Fear of movement was assessed with the valid and reliable shortened version of the Tampa Scale of Kinesiophobia (TSK-11).\textsuperscript{31} The TSK-11 consists of 11 statements with four answer options varying from ‘strongly disagree’ to ‘strongly agree’. The total score ranges from 11 to 44 points, with a higher score reflecting greater fear of movement.

**Follow-up**

Five days after filling-in the baseline questionnaire (T1) patients received a follow-up questionnaire consisting of the Dutch STarT MSK tool, the NPRS to assess the average pain in the past week and the Global Perceived Effect (GPE) scale\textsuperscript{32} to measure recovery. The GPE scale consists of the question ‘To what extend have your complaints improved since filling-in the baseline questionnaire?’, scored on a 7-point Likert scale (1 = completely recovered; 7 = worse than ever).
Main study parameters

The measurement properties construct validity and test-retest reliability of the Dutch STarT MSK tool were evaluated according to the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) checklist.\textsuperscript{33,34}

Construct validity

To investigate whether the specific items of the Dutch STarT MSK tool measure the intended biomedical and psychosocial prognostic factors, we assessed the construct validity by comparing scores of the separate items with reference questionnaires. This method was used because the STarT MSK tool is a formative model,\textsuperscript{35} which means that each item contributes a part of the construct and together they will give a prognosis for persisting musculoskeletal-related disability. \textit{A priori} we expected a moderate to high positive correlation ($r \geq 0.3$) between STarT MSK item 1 with the NPRS, item 6 with the timeline question of the Brief IPQ, item 7 with the single-item question on ‘comorbidity’, item 8 with the HADS-D, item 9 with the TSK-11, item 10 with the single-item question on ‘pain duration’; and a moderate to high

---

Figure 1. Graphical representation of the STarT MSK tool.
negative correlation ($r \leq -0.3$) between item 2 with the PSEQ and item 4 with the SF-36 PF, based on the comparability of the domains being measured. We expected a moderate to high positive correlation ($r \geq 0.3$) between the STarT MSK bothersomeness item 3 with the NPRS and a moderate to high negative correlation ($r \leq -0.3$) between item 3 with the SF-36 PF, as bothersomeness has been associated with pain and disability.\cite{36} Finally, we expected a low positive correlation ($r < 0.3$) between item 5 and the NPRS, as this item focuses on the location of pain and not on pain intensity.

**Test-retest reliability**

To investigate whether the scores of the Dutch STarT MSK tool are consistent over time, we evaluated the agreement between scores on baseline and after 5 days. The time interval was considered long enough to prevent for recall bias, given the large number of questionnaires patients had to complete at baseline. Next, 5 days were considered short enough to prevent substantial improvement.\cite{37} The test-retest sample comprised patients who completed the Dutch STarT MSK tool at T0 and T1. Furthermore, a sensitivity analysis was performed in a subset of patients reporting stable musculoskeletal pain symptoms during the test-retest period. In concordance with the criteria proposed by Bier et al.,\cite{14,38} patients were considered stable between T0 and T1 when they scored ‘slightly worsened’, ‘no change’, or ‘slightly improved’ on the GPE, and had a stable pain score (i.e., the same score on the NPRS plus or minus one point compared with T0).

**Sample size**

To evaluate construct validity and test-retest reliability, a minimal sample size of 50 patients is advised and a sample size of 100 patients is adequate according to the COSMIN checklist.\cite{37,39} In this study, we aimed for an adequate sample size of at least 100 patients.\cite{39}

**Statistical analysis**

Statistical analysis was performed using IBM SPSS Statistics version 23.0 (Armonk, New York, USA). Descriptive statistics were calculated for baseline characteristics of the study population, with continuous variables presented using mean and standard deviations. Categorical and nominal/dichotomous data were presented as proportions for each category.

**Construct validity**

Pearson’s correlations were calculated between specific items of the Dutch STarT MSK tool and their reference questionnaires. The construct validity was defined as good if at least 75% of the *a priori* hypotheses could be confirmed.\cite{37}

**Test-retest reliability**

The quadratic-weighted kappa was calculated for overall scores of the Dutch STarT MSK tool and prognostic subgroups, with $\kappa \leq 0$ indicating poor agreement, $\kappa = 0.01-0.20$ slight agreement, $\kappa = 0.21-0.40$ fair agreement, $\kappa = 0.41-0.60$ moderate agreement, $\kappa = 0.61-0.80$ substantial agreement and $\kappa = 0.81-1.00$ almost perfect agreement.\cite{40}
RESULTS

In total, 44 physiotherapists were instructed and 22 physiotherapists from 11 primary care physiotherapy clinics actually registered patients. Of the participating physiotherapists, the majority was specialised as manual therapist or sports physiotherapist. A total of 167 patients were registered, of whom 146 patients were included (Figure 2). Four patients were excluded from the analysis, because they did not fully complete the baseline questionnaire. For the follow-up questionnaires, a 100% follow-up rate was achieved.

Figure 2. Flowchart for inclusion of patients.

Baseline characteristics of the study population are presented in Table 1. The mean age of study participants was 48.2 ± 15.6 years (range 18 – 81), and 57.0% were female. The study population consisted of 142 patients with musculoskeletal pain, of whom 44 patients (31.0%) had back pain, 35 (24.7%) shoulder pain, 28 (19.7%) multisite pain, 25 (17.6%) knee pain and 10 (7.0%) neck pain. Of the patients with multisite pain, the majority (57.1%) reported neck and shoulder pain. At baseline, 74 patients (52.1%) were categorised as low risk, 64 (45.1%) as medium risk and 4 (2.8%) as high risk. Age and gender were comparable across risk groups. For each increase in risk profile, there was a corresponding increase in pain intensity, disability, timeline illness perception, depression and fear of movement.
### Table 1. Baseline characteristics of the study population.*

<table>
<thead>
<tr>
<th></th>
<th>Total (n = 142)</th>
<th>Low risk (n = 74)</th>
<th>Medium risk (n = 64)</th>
<th>High risk (n = 4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>81 (57.0)</td>
<td>40 (54.1)</td>
<td>39 (60.9)</td>
<td>2 (50.0)</td>
</tr>
<tr>
<td>Age in years</td>
<td>48.2 ± 15.6</td>
<td>47.9 ± 16.8</td>
<td>47.9 ± 14.5</td>
<td>56.8 ± 8.2</td>
</tr>
<tr>
<td>Pain site</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neck</td>
<td>10 (7.0)</td>
<td>6 (8.1)</td>
<td>3 (4.7)</td>
<td>1 (25.0)</td>
</tr>
<tr>
<td>Back</td>
<td>44 (31.0)</td>
<td>22 (29.7)</td>
<td>21 (32.8)</td>
<td>1 (25.0)</td>
</tr>
<tr>
<td>Shoulder</td>
<td>35 (24.7)</td>
<td>23 (31.1)</td>
<td>12 (18.8)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Knee</td>
<td>25 (17.6)</td>
<td>12 (16.2)</td>
<td>12 (18.8)</td>
<td>1 (25.0)</td>
</tr>
<tr>
<td>Multisite</td>
<td>28 (19.7)</td>
<td>11 (14.9)</td>
<td>16 (25.0)</td>
<td>1 (25.0)</td>
</tr>
<tr>
<td>Episode duration</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;3 months</td>
<td>72 (50.7)</td>
<td>46 (62.2)</td>
<td>25 (39.1)</td>
<td>1 (25.0)</td>
</tr>
<tr>
<td>3 to 6 months</td>
<td>26 (18.3)</td>
<td>16 (21.6)</td>
<td>9 (14.1)</td>
<td>1 (25.0)</td>
</tr>
<tr>
<td>6 months or longer</td>
<td>44 (31.0)</td>
<td>12 (16.2)</td>
<td>30 (46.9)</td>
<td>2 (50.0)</td>
</tr>
<tr>
<td>Presence of comorbidity</td>
<td>59 (41.5)</td>
<td>28 (37.8)</td>
<td>29 (45.3)</td>
<td>2 (50.0)</td>
</tr>
<tr>
<td>Dutch STarT MSK tool score†</td>
<td>4.2 ± 2.3</td>
<td>2.4 ± 1.2</td>
<td>6.0 ± 1.0</td>
<td>9.5 ± 0.6</td>
</tr>
<tr>
<td>Dutch STarT MSK tool risk profile</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>74 (52.1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medium</td>
<td>64 (45.1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>4 (2.8)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pain intensity§‡</td>
<td>5.4 ± 2.0</td>
<td>4.5 ± 2.0</td>
<td>6.2 ± 1.6</td>
<td>8.0 ± 0.0</td>
</tr>
<tr>
<td>Mild (0 – 5)</td>
<td>62 (43.7)</td>
<td>47 (63.5)</td>
<td>15 (23.4)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Moderate (6 – 7)</td>
<td>60 (42.3)</td>
<td>21 (28.4)</td>
<td>39 (60.9)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Severe (8 – 10)</td>
<td>20 (14.1)</td>
<td>6 (8.1)</td>
<td>10 (15.6)</td>
<td>4 (100.0)</td>
</tr>
<tr>
<td>Pain self-efficacy§</td>
<td>49.4 ± 9.6</td>
<td>53.5 ± 7.3</td>
<td>45.6 ± 9.4</td>
<td>34.0 ± 9.9</td>
</tr>
<tr>
<td>Disability§</td>
<td>75.1 ± 19.5</td>
<td>83.0 ± 13.7</td>
<td>67.2 ± 21.4</td>
<td>56.3 ± 19.7</td>
</tr>
<tr>
<td>Timeline illness perception§</td>
<td>4.2 ± 2.9</td>
<td>3.1 ± 2.2</td>
<td>5.3 ± 3.0</td>
<td>7.5 ± 2.1</td>
</tr>
<tr>
<td>Depression§</td>
<td>2.3 ± 2.5</td>
<td>1.3 ± 1.8</td>
<td>3.1 ± 2.4</td>
<td>7.3 ± 4.0</td>
</tr>
<tr>
<td>Fear of movement§</td>
<td>20.0 ± 5.6</td>
<td>18.5 ± 5.1</td>
<td>21.2 ± 5.4</td>
<td>28.3 ± 7.2</td>
</tr>
</tbody>
</table>

* Values are numbers (percentage) or mean ± standard deviation.
† Dutch STarT MSK tool score (0 – 12); Pain intensity (0 – 10); Pain self-efficacy (0 – 60); Disability (0 – 100); Timeline illness perception (0 – 10); Depression (0 – 21); Fear of movement (11 – 44).

#### Construct validity

The Pearson’s correlations between the separate items of the Dutch STarT MSK tool and their reference questionnaires are presented in Table 2. The highest correlations were found between item 1 with the NPRS and item 10 with the single-item question on ‘pain duration’. The correlations between item 2 with the PSEQ and item 9 with the TSK-11 were lower than hypothesised. Of the predefined hypotheses, 81.8% were confirmed which indicates a good construct validity.
Table 2. Pearson’s correlation between the Dutch STarT MSK tool and reference questionnaires.

<table>
<thead>
<tr>
<th>STarT MSK item</th>
<th>Reference questionnaire</th>
<th>A priori</th>
<th>r</th>
<th>Correlation</th>
<th>Expected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item 1</td>
<td>NPRS</td>
<td>r ≥ 0.3</td>
<td>0.815</td>
<td>High</td>
<td>Yes</td>
</tr>
<tr>
<td>Item 2</td>
<td>PSEQ</td>
<td>r ≤ -0.3</td>
<td>-0.221</td>
<td>Low</td>
<td>No</td>
</tr>
<tr>
<td>Item 3</td>
<td>NPRS</td>
<td>r ≥ 0.3</td>
<td>0.480</td>
<td>Moderate</td>
<td>Yes</td>
</tr>
<tr>
<td>Item 3</td>
<td>SF-36 PF</td>
<td>r ≤ -0.3</td>
<td>-0.417</td>
<td>Moderate</td>
<td>Yes</td>
</tr>
<tr>
<td>Item 4</td>
<td>SF-36 PF</td>
<td>r ≤ -0.3</td>
<td>-0.466</td>
<td>Moderate</td>
<td>Yes</td>
</tr>
<tr>
<td>Item 5</td>
<td>NPRS</td>
<td>r &lt; 0.3</td>
<td>0.125</td>
<td>Low</td>
<td>Yes</td>
</tr>
<tr>
<td>Item 6</td>
<td>Timeline question Brief IPQ</td>
<td>r ≥ 0.3</td>
<td>0.570</td>
<td>High</td>
<td>Yes</td>
</tr>
<tr>
<td>Item 7</td>
<td>Single item ‘comorbidity’</td>
<td>r ≥ 0.3</td>
<td>0.344</td>
<td>Moderate</td>
<td>Yes</td>
</tr>
<tr>
<td>Item 8</td>
<td>HADS-D</td>
<td>r ≥ 0.3</td>
<td>0.393</td>
<td>Moderate</td>
<td>Yes</td>
</tr>
<tr>
<td>Item 9</td>
<td>TSK-11</td>
<td>r ≥ 0.3</td>
<td>0.294</td>
<td>Low</td>
<td>No</td>
</tr>
<tr>
<td>Item 10</td>
<td>Single item ‘pain duration’</td>
<td>r ≥ 0.3</td>
<td>0.795</td>
<td>High</td>
<td>Yes</td>
</tr>
</tbody>
</table>

r = Pearson’s correlation.
NPRS = Numeric Pain Rating Scale, PSEQ = Pain Self-Efficacy Questionnaire. SF-36 PF = Physical Functioning subscale of the 36-item Short Form Health Survey, IPQ = Illness Perception Questionnaire, HADS-D = depression subscale of the Hospital Anxiety and Depression Scale, TSK-11 = 11-item Tampa Scale of Kinesiophobia.

Test-retest reliability
In total, 142 patients completed the Dutch STarT MSK tool at T0 and T1, of whom 77 patients were regarded as stable. On average, the time between T0 and T1 was 7 days. The quadratic-weighted kappa coefficients for the overall tool scores and prognostic subgroups were 0.707 and 0.653, respectively, indicating substantial agreement (Table 3). Distribution of prognostic subgroups was skewed due to the very low prevalence of patients at high risk. Test-retest reliability increased to 0.746 for the overall tool scores and decreased to 0.603 for prognostic subgroups, when agreement was calculated in the subset of 77 patients reporting stable musculoskeletal pain symptoms.

Table 3. Quadratic-weighted kappa for the Dutch STarT MSK tool.*

<table>
<thead>
<tr>
<th></th>
<th>Test-retest sample (n = 142)</th>
<th>Subset of patients reporting stable symptoms (n = 77)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall tool scores</td>
<td>0.707 (0.623-0.792)</td>
<td>0.746 (0.658-0.834)</td>
</tr>
<tr>
<td>Prognostic subgroups</td>
<td>0.653 (0.543-0.762)</td>
<td>0.603 (0.452-0.755)</td>
</tr>
</tbody>
</table>

* Values are represented as quadratic-weighted kappa coefficients with 95% confidence intervals.
κ ≤ 0: poor agreement, κ = 0.01-0.20: slight agreement, κ = 0.21-0.40: fair agreement, κ = 0.41-0.60: moderate agreement, κ = 0.61-0.80: substantial agreement, κ = 0.81-1.00: almost perfect agreement.

DISCUSSION
This is the first study evaluating the construct validity and test-retest reliability of the Dutch STarT MSK tool in patients with musculoskeletal pain in primary care physiotherapy. The results of this validation study showed a good construct validity of the Dutch STarT MSK tool, as 81.8% of the predefined hypotheses were confirmed. The test-retest reliability of the Dutch STarT MSK tool was substantial for the overall tool scores and prognostic subgroups in the
test-retest sample, with quadratic-weighted kappa coefficients of 0.707 and 0.653, respectively. Test-retest reliability remained substantial for the overall scores and decreased slightly to a kappa of 0.603 for prognostic subgroups in the subset of patients reporting stable symptoms between T0 and T1.

For the construct validity of the Dutch STaRT MSK tool, nine of the eleven predefined hypotheses were confirmed, which suggests that nearly all items measure the intended biomedical and psychosocial prognostic factors. These results are in line with the Dutch SBT for patients with low-back pain and the modified SBT for patients with neck pain, both indicating good construct validity. Since the validation study of the original English STaRT MSK tool assessed the construct validity by comparing risk subgrouping based on the STaRT MSK tool with risk subgrouping based on the Örebro Musculoskeletal Pain Screening Questionnaire (ÖMPSQ), the results found in the current study cannot be compared with the original STaRT MSK tool. We refrained from the method used in the UK study, because the Dutch version of the ÖMPSQ is only validated in patients with low-back pain.

As the STaRT MSK tool is an instrument with predictive purposes, it is important to classify clinically stable individuals into the same prognostic subgroup over time, otherwise it can influence the predictive ability of the tool. For prognostic subgroups, the kappa coefficient decreased from 0.653 (test-retest sample) to 0.603, when agreement was calculated using the clinically stable patients. The kappa of 0.603 is just below the cut-off point of 0.610 for substantial agreement and theoretically should be interpreted as moderate agreement. However, distribution of prognostic subgroups was skewed due to the very low prevalence of patients at high risk, which increases chance agreement and reduces the value of kappa accordingly. Therefore, it seems more appropriate to consider the test-retest reliability of prognostic subgroups as substantial. The test-retest reliability of the Dutch STaRT MSK tool is comparable with the corresponding quadratic-weighted kappa coefficients of the original SBT and translated versions (0.58 to 0.79).

In the present study, only 4 patients (2.8%) were categorised as high risk. This proportion of high risk patients is lower compared to the UK study, as well as the initial SBT validation study and translated versions (9.8% to 23.2%). Several factors might have influenced the distribution in risk profiles found in our study. First, the setting in which patients were included might explain the skewed distribution. Most of the participating physiotherapists were specialised as manual therapist or sports physiotherapist. In this way, it is possible that the more severe psychosomatic cases and potentially high risk patients were missed, since they probably visit specialised psychosomatic physiotherapists. Second, for feasibility reasons, patients received the baseline questionnaire after the first consultation. Especially in acute high risk patients, psychosocial risk factors might have been addressed during the first consultation, which could influence the results. In a study regarding changes in SBT categorisation, 81.8% of the high risk patients were categorised differently within a few weeks. Therefore, in the current study it is possible that some acute high risk patients might have shifted from high to medium risk before filling-in the baseline questionnaire.
The strength of this study is that we evaluated the measurement properties of the Dutch STarT MSK tool in primary care physiotherapy. In the Netherlands, physiotherapists in particular are involved in the treatment of patients with musculoskeletal pain and therefore are likely the primary users of the tool. Another strength is that we achieved an appropriate sample size (>100 patients) and 100% follow-up rate by e-mailing and phone calling non-responders repeatedly when necessary. The present study has some limitations. First, there was a lack of participating specialised psychosomatic physiotherapists, which might resulted in the low proportion of high risk patients. Second, we did not perform additional analyses across pain sites because therefore a larger sample size is required, so it is not clear if there are any differences between the regional musculoskeletal pain presentations. However, in a preliminary validation study of the original STarT MSK tool, results were comparable across pain sites. Finally, we could not compare our results with the recently refined English STarT MSK tool, because the manuscript on the validation of this tool is in preparation for publication.

This validation study demonstrates promising results regarding the measurement properties of the Dutch STarT MSK tool. However, further psychometric evaluation is required before the tool can be implemented in clinical practice. First, the predictive validity of the tool has to be established. We are currently evaluating this in a prospective cohort study while applying usual care as treatment. Second, additional analyses across pain sites are recommended to investigate whether the results are comparable across the regional musculoskeletal pain presentations. For these additional analyses, a minimal sample size of 50 patients for each pain site is advised. Finally, the Dutch STarT MSK tool is not just a prognostic screening tool, but can be used to stratify patients for the appropriate matched treatment as well. Currently, there are only matched treatment packages available for patients with low-back pain. Therefore, future research should also focus on developing stratified care approaches for patients with neck, shoulder, knee and multisite pain.

CONCLUSION

This study is a first step toward identifying whether the Dutch STarT MSK tool could be useful for clinical practice. According to the psychometric analysis, the tool showed a good construct validity and substantial test-retest reliability in patients with musculoskeletal pain in primary care physiotherapy. However, based on these measurement properties it is not clear yet if the Dutch STarT MSK tool is a valid instrument for allocating patients with musculoskeletal pain into prognostic subgroups. Therefore, further research is underway to evaluate the predictive validity of the tool.
REFERENCES


Assessment. 2012.


40. Landis JR and Koch GG. The measurement of observer agreement for categorical data.
Biometrics. 1977; 33: 159-74.


APPENDIX 1

The Keele STarT MSK Tool©

Patient name: ________________________________ Date: ______________

For questions 1-9, think about just the last two weeks:

**Pain intensity**

1. On average, how intense was your pain [where 0 is “no pain” and 10 is “pain as bad as it could be”]?  

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

Please cross one box for each question below

<table>
<thead>
<tr>
<th></th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2. Do you often feel unsure about how to manage your pain condition?  

3. Over the last two weeks, have you been bothered a lot by your pain?  

4. Have you only been able to walk short distances because of your pain?  

5. Have you had troublesome joint or muscle pain in more than one part of your body?  

6. Do you think your condition will last a long time?  

7. Do you have other important health problems?  

8. Has pain made you feel down or depressed in the last two weeks?  

9. Do you feel it is unsafe for a person with a condition like yours to be physically active?  

10. Have you had your current pain problem for 6 months or more?  

**Total score:**  

<table>
<thead>
<tr>
<th>Total score:</th>
<th>0-4 = Low risk</th>
<th>5-8 = Medium risk</th>
<th>9-12 = High risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
</tbody>
</table>
## APPENDIX 2

**Dutch version of the STarT MSK tool**

Naam: ___________________________ Datum: _____________

Denk bij het beantwoorden van de vragen 1-9 alleen aan de laatste 2 weken:

### Pijnintensiteit

1. Gemiddeld genomen, hoe hevig was uw pijn (waarbij 0 betekent “geen pijn” en 10 “ergste pijn denkbaar”)?

<p>| | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

Kruis alstublieft één vakje aan bij elke onderstaande vraag

<table>
<thead>
<tr>
<th>Ja</th>
<th>Nee</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

2. Voelt u zich vaak onzeker over hoe u met uw pijn moet omgaan? ☐ ☐

3. Bent u in de laatste 2 weken veel gehinderd door uw pijn? ☐ ☐

4. Bent u door uw pijn alleen in staat geweest korte afstanden te lopen? ☐ ☐

5. Heeft u hinderlijke gewrichts- of spierpijn gehad in meer dan één lichaamsdeel? ☐ ☐

6. Denkt u dat uw klacht lang zal aanhouden? ☐ ☐

7. Heeft u andere belangrijke gezondheidsproblemen? ☐ ☐

8. Heeft u zich in de laatste 2 weken somber of depressief gevoeld door uw pijn? ☐ ☐

9. Heeft u het gevoel dat het voor iemand met uw klacht onveilig is om lichamelijk actief te zijn? ☐ ☐

10. Heeft u uw huidige pijnklacht sinds 6 maanden of langer? ☐ ☐

### Totaalscore:

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0-4 = Laag risico</td>
<td>5-8 = Gemiddeld risico</td>
<td>9-12 = Hoog risico</td>
</tr>
<tr>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

[van den Broek A.G.] [Construct validity and test-retest reliability of the Dutch STarT MSK tool] 21
SAMENVATTING

Achtergrond De STarT MSK tool is een recent in Engeland ontwikkeld en gevalideerd prognostisch screeningsinstrument op basis waarvan patiënten met musculoskeletale pijn worden ingedeeld in drie risicogroepen. Om te bepalen of de STarT MSK tool ook in Nederland van toegevoegde waarde kan zijn, moeten eerst de meeteigenschappen worden onderzocht.

Doelstelling Het bepalen van de construct validiteit en test-hertest betrouwbaarheid van de Nederlandse STarT MSK tool bij patiënten met musculoskeletale pijn in de fysiotherapeutische eerstelijnszorg.

Methode Fysiotherapeuten includeerden patiënten met musculoskeletale pijn in de leeftijd van 18 jaar of ouder. Patiënten vulden een vragenlijst in op baseline en na 5 dagen. De construct validiteit werd bepaald door het vergelijken van scores van de afzonderlijke items van de STarT MSK tool met referentie vragenlijsten. A priori hypotheses werden getoetst door het berekenen van Pearson’s correlatie coëfficiënten. De test-hertest betrouwbaarheid werd bepaald door het berekenen van kwadratisch gewogen kappa coëfficiënten voor de totaalscore en prognostische subgroepen.

Resultaten In totaal werden 142 patiënten geïncludeerd voor de psychometrische analyse. Van de 142 patiënten werden 74 patiënten (52,1%) gecategoriseerd als laag risico, 64 (45,1%) als gemiddeld risico en 4 (2,8%) als hoog risico. Voor construct validiteit werden negen van de elf hypotheses bevestigd. De correlaties tussen item 2 (pijn self-efficacy) en item 9 (bewegingsangst) met de desbetreffende referentie vragenlijsten waren lager dan verwacht. De kwadratisch gewogen kappa coëfficiënten voor de totaalscore en prognostische subgroepen waren respectievelijk 0,707 en 0,653. Test-hertest betrouwbaarheid steeg naar 0,746 voor de totaalscore en daalde naar 0,603 voor prognostische subgroepen, wanneer overeenstemmingsindex berekend bij de 72 patiënten met stabiele symptomen.

Conclusie De Nederlandse STarT MSK tool laat een goede construct validiteit en voldoende tot goede test-hertest betrouwbaarheid zien bij patiënten met musculoskeletale pijn in de fysiotherapeutische eerstelijnszorg.

Klinische relevantie Deze validatiestudie laat veelbelovende resultaten zien ten aanzien van de meeteigenschappen van de STarT MSK tool. Echter, verdere psychometrische analyse met betrekking tot de predictieve validiteit en aanvullende analyses voor de verschillende pijnregio’s zijn vereist voordat de tool kan worden geïmplementeerd in de klinische praktijk.

Trefwoorden Musculoskeletale pijn, fysiotherapie, prognostisch screeningsinstrument, validiteit, betrouwbaarheid.