

Elsevier required licence: © <2017>. This manuscript version is made available under the CC-BY-NC-ND 4.0 license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Title:

Validity and reproducibility of the modified STarT Back Tool (Dutch version) for patients with neck pain in primary care

Authors:

Jasper D. Bier^{1,2}

Raymond W. J. G. Ostelo^{3,4}

Bart W. Koes¹

Arianne P. Verhagen¹

1. Department of General Practice, Erasmus MC, PO Box 2040, 3000 CA, Rotterdam, The Netherlands
2. Fysiotherapie Fascinatio, Capelle aan den IJssel, The Netherlands
3. Department of Epidemiology and Biostatistics, VU University Medical Centre Amsterdam and the EMGO Institute for Health and Care Research, The Netherlands
4. Department of Health Sciences, Faculty of Earth and Life Sciences, VU University Amsterdam, The Netherlands.

Correspondence to:

Jasper D. Bier, Department of General Practice, Erasmus MC, PO Box 2040, 3000CA Rotterdam, The Netherlands

T +31 6 26536787 **E** j.bier@erasmusmc.nl **URL** www.erasmusmc.nl

Total number of pages:

Manuscript = 18 pages (including title page, abstract, figure / table legends and references)

Figures = 1

Tables = 5

Word count: abstract: 300

Word count article (including abstract) 3453

ABSTRACT

Objective: To evaluate the reliability and validity of the Dutch Version of the STarT Back screening Tool (SBT) for patients with neck pain.

Methods: We modified the original SBT for back pain to be used in patients with neck pain. General practitioners and physiotherapists included patients who completed a baseline questionnaire and follow-up measurements at three days and three months. The construct validity was assessed by formulating hypotheses about the expected magnitude of the Pearson's correlation between SBT and other questionnaires measuring pain, activity, kinesiophobia and catastrophizing. The reproducibility was assessed in the first week using test-retest measurements and by calculating the quadratic weighted kappa and the specific agreement. Predictive validity was assessed using relative risk ratios (RR) for, amongst others, persisting disability at three months. Content validity was analysed using floor- and ceiling effects.

Results: In total 100 patients were included, 58.0% was categorized as "low-risk" for poor outcome, 37.0% as "medium-risk" and 5.0% as "high-risk". For the construct validity we found, as expected, a moderate to high correlation for all questions except for activity question 3 (lower than expected correlation) with their reference questionnaires. The reproducibility had a quadratic weighted kappa of 0.85 and the specific agreement of 90.9% for "low-risk" and 66.7% for "medium-risk".

The RR for persisting disability and persisting pain for "medium-risk" compared to "low risk" were **1.5 (95% C.I. 0.9 – 2.4) and 1.0 (95% C.I. 0.7-1.5)** respectively. For "high-risk", compared to "low-risk" **RR=1.5 (95% C.I. 0.5-4.1) and 1.2 (95% C.I. 0.5-2.7)** respectively. We found no floor and ceiling effects.

Conclusion: The original SBT for back pain is successfully modified **to fit** patients with neck pain in Dutch primary care. The psychometric analysis showed sufficiently reliable outcomes, although **the predictive validity in primary care is limited.**

Keywords:

GP, physical therapy, neck pain, classification, validation, STarT Back Tool

BACKGROUND

Globally, neck pain is the 4th largest musculoskeletal disorder causing disability.¹ Numbers from the Netherlands institute for health service research (NIVEL) found that in 2015 neck pain is the most prevalent disorder in the Dutch physical therapy.² The estimated one-year incidence of neck pain varies between 10.4 to 21.3%.³ In patients with acute neck pain, the pain and disability decrease in the first six weeks with about 45% but little or no decrease can be found afterwards.⁴ A Dutch cohort found that after one year 76% of the patients with neck pain reported to be fully recovered or much improved, indicating that still in many patients the complaints persist over time and or are recurrent⁵ The annual cost of neck pain in the Netherlands was estimated to be \$668 million in 1996, unfortunately more recent data for The Netherlands are not available.⁶

The Neck Pain Task Force (NPTF) and the Dutch physiotherapists' guideline on neck pain classify neck pain into four grades based on interference with activities of daily living, presence of neurologic signs (grades I-III) or signs and symptoms of major structural pathology (grade IV).^{7,8} A classification based on the course of the pain (normal or delayed recovery) often determines whether physiotherapeutic treatment is indicated, as when the course of the pain is normal, treatment is often not indicated. Subgrouping patients is becoming an increasingly popular method for applying targeted treatment. It has the potential to optimise treatment benefits and maximise healthcare efficiency. For low back pain in primary care the Keele STarT Back Tool (Subgroups for Targeted Treatment) (SBT) is probably the best known tool for subgrouping back pain patients combined with a targeted treatment. The SBT focuses on the combination of limitations in patients' activity, patients' pain as well as several psychosocial factors known to influence patients' recovery. It is developed to allocate primary care patients with low back pain into three subgroups concerning their prognosis: low, moderate or high risk for persisting disability and to apply the appropriate stratified care.^{9,10} The SBT consists of nine questions, eight true/false questions and one question with a 5-point Likert-scale as answer option. Based on these independent factors it aims to predict poor disability with each factor adding to the likelihood of a poor prognosis. For each subgroup a targeted treatment is advised. In short; low risk patients receive information and advice, in addition medium risk patients receive standardised physiotherapy to address symptoms and function and high-risk patients receive, in addition, psychologically informed physiotherapy to address physical symptoms and function, and also psychosocial obstacles to recovery.¹¹ The SBT has been and validated for low back pain in the UK⁹ and translated in several other languages.¹²⁻¹⁹ A few preliminary studies are performed for other musculoskeletal pain conditions such as lumbar stenosis, knee pain, shoulder pain and neck pain but these are not ready for clinical implementation.²⁰⁻²² Our aim is to translate and modify the SBT for patients with neck pain and to evaluate the validity and reliability of the modified SBT (SBT-Neck) in Dutch primary care.

METHOD

Developing the SBT-Neck

The Dutch SBT for low back pain was used as a basis for the neck version. Initially it was one combined SBT tool for both patients with neck and low back pain.¹⁹ Next, we performed a preliminary field-test with two general practitioners (GP's) and one physiotherapist (PT) working in primary care to analyse the instruments' feasibility in 140 patients with neck pain or low back pain of any duration. Out of the 140 patients 24.3% had neck pain, 42.1% had back pain and 32.9% had both. We found that especially patients who suffered from both neck and lower back pain (n=46) had problems answering the questions. They could not distinguish well between low back pain and neck pain.

During a subsequent expert meeting we decided to return to the initial low back pain version and develop a separate neck pain version. In most questions "back" was replaced with "neck" and radiating pain in the legs was modified as radiating pain in the arms. Question 3 was altered into "I have used my arms and neck less due to my neck pain" instead of "I have only walked short distances because of my back pain". The changes were based on consensus in the working group. (Appendix 1 (in Dutch)). Appendix 2 shows the comparison between the original SBT, the Dutch neck pain version and its English translation.

Design

We performed a clinimetric substudy as part of a prospective cohort (PRINS study; Prevalence of Risk groups in neck- and back pain patients according to the STarT back screening tool).¹⁹ All patients that consulted primary care for low back or neck pain were asked to answer baseline and follow-up questionnaires and received regular care by their clinician. The study was approved by the medical ethics committee of the Erasmus University, Rotterdam, The Netherlands. (MEC-2014-256). For this study we only use the data of the patients with neck pain of the PRINS-cohort.

Participants

Clinicians

We invited clinicians to participate in this study who had previously shown their interest in evaluating the SBT. They all attended a meeting in which the study was explained and they received the study protocol. They also received posters, information-brochures and informed consent forms for the patients.

Patients

In November 2014 we started including patients until May 2015. When a patient consulted a GP or PT thru referral or direct access for their neck pain they were asked to participate in the PRINS study. They were included when they had non-specific neck pain (grade I-III) and when the patient was 18 years or older, could speak, read and write in Dutch and had an email address. Patients were excluded if during the consultation the GP or PT found red flags indicating a possible specific underlying pathology (grade IV) e.g. infection, fracture, cauda equina or tumor. Patients were given oral and written information about the procedure of data collection and the aim of the study. When a patient was willing to participate they signed an informed consent, handed it back to their clinician, which registered them online. The patient immediately received an email with a link to the baseline questionnaire.

Treatment

The patients received usual care by their clinician. The clinician was unaware of the results of the baseline questionnaire including the SBT-score.

Measurements

Baseline

At baseline (T0) patients filled out a questionnaire consisting of demographic variables (such as age, gender) and the SBT-Neck. We measured the average pain in the past week using the 11-point Numeric Pain Rating Scale (NPRS)²⁴ ranging from 0="no pain" to 10="worst imaginable pain". Disability was assessed using the Neck Disability Index (NDI)²⁵ consisting of 10 statements with a 6-point scale ranging from 0="no limited" to 5="completely limited". The score is doubled to gaining a total score ranging from 0 to 100, with a higher score indicating more disability. We measured fear of movement/(re)injury using the Tampa Scale of Kinesiophobia (TSK)²⁶ consisting of 17 statements with four answer options varying from 1="highly disagree" to 4="highly agree". The total score ranges from 17 to 68, a higher score indicating a higher level of kinesiophobia. To assess the level of catastrophizing we used the Pain Catastrophizing Scale (PCS), which consists of 13 statements with each a 5-point scale answers option ranging from 0="not at all" to 4="always".²⁷ The total score ranges from 0 to 65, a higher score indicating a higher level of catastrophizing. Finally we assessed quality of life using the EQ-5D²⁸ consisting of six questions. The first five questions have a 3-point Likert scale answer options ranging from 1="no problems" to 3="severe problems" and the sixth question is a health status question ranging from 0="worst imaginable health" to 100="best imaginable health".

Follow-up

Three days after inclusion (T1) a follow-up questionnaire was sent in order to investigate the reliability of the SBT-Neck. It consisted of the SBT-Neck, the NPRS and the General Perceived Effect scale (GPE) to assess pain and recovery respectively. The answer options on the GPE range from 1="fully recovered" to 7="worse than ever". We considered three days short enough to prevent substantial improvement and long enough, in combination with all other questionnaires at baseline, to reduce recall bias.²³ For practical reasons the test-retest questionnaire was added only for the patients that were included during the last three months of the inclusion period.

Three months after inclusion (T2), the patients received a follow-up questionnaire consisting of the GPE and NDI. At the same time we sent a questionnaire to the GP to inquire about the number of visits, prescribed medication, referrals to physiotherapists or medical professionals and requested diagnostic imaging and blood tests. We sent a similar questionnaire to the PT to inquire about treatment data such as date of first and last treatment, number of treatment sessions, questionnaires used and the aim and means of treatment. All questionnaires were handled and stored through Lime Survey 2.05.

Sample size

The minimum sample size of 50 persons for all aspects of the clinimetric study is advised by Terwee et al; we aimed at a minimum of 100 persons.²³

Statistical analysis

We analysed the data to describe patients' characteristics, expressed using frequencies, means and standard deviations. The risk profiles distribution and their characteristics are reported.

For the construct validity we first analyzed the characteristics across the SBT risk profile to determine the discriminant validity. Next, we calculated the Pearson's correlation coefficient for each item of the SBT-Neck and their reference questionnaires based on the comparability of the domains of measurement.^{29,30} We expected a moderate ($r \geq 0.3, < 0.5$) to high ($r \geq 0.5$) correlation between the SBT-Neck questions 1 with single item question 'referred pain', activity-item 3 and 4 with the NDI, kinesiophobia-question 5 with the TSK, catastrophizing-questions 6 and 7 with the PCS and the bothersome-question 9 with the NPRS. We included no reference questionnaire for question 2 and no questionnaire to measure depression (question 8).

For the reproducibility we selected the patients that remained stable between baseline (T0) and T1. Patients were considered stable when they scored "slightly improved", "no change" or "slightly worsened" on the GPE at T1. As there is some doubt in the literature whether the GPE actually can detect change, we combined the stable GPE score with a stable pain score meaning the NPRS on T1; meaning the same score plus or minus one point compared to the baseline score.³¹ We then calculated the quadratic weighted kappa for the ability to distinguish between groups and the specific agreement for the agreement within a group. The kappa will be interpreted as ≤ 0 = poor agreement; .01-.20 = slight; .21-.40 = fair; .41-.60 = moderate; .61-.80 = substantial and .81-1 = almost perfect agreement.³² The specific agreement is calculated for each risk profile separately.²⁹ For example, patients who are "low-risk" on T0 and T1 are calculated as a proportion of patients that were "low-risk" on either of the two measurements. In collaboration with de Vet we modified the specific agreement to fit a 3x3 table as shown in table 1 because the original method is done in a 2x2 table.

We determined the predictive validity by reporting the relative Risk Ratio (RR) for "medium-risk" and "high-risk", both compared to "low-risk" in their ability to predict the outcome at three months. The outcomes are: 1) Persisting disability, defined as the number of individuals with an NDI that decreased less than the minimally clinical important difference (MCID) (14 points, range 0-100)³³ compared to the baseline NDI score, 2) Persisting pain, defined as number of individuals with an NPRS that decreased less than the MCID (2 points, range 0-10)³⁴ compared to the baseline NPRS score and 3) Perceived recovery, defined as either "completely recovered" or "much improved" on the GPE.³⁵

Limited content validity is indicated by the presence of more than 15 percent of the patients reached either the floor (0/9 points) or ceiling (9/9 points) effects of the SBT.²³

RESULTS

In total, 12 GPs and 33 PTs included patients in the PRINS-study. They included 100 patients with neck pain. Loss to follow up at three months was 7 (7.0%) (figure 1). At baseline we found that 58 (58.0%) patients were categorized as "low-risk", 37 (37.0%) as "medium-risk" and 5 (5.0%) as "high-risk" (table 2). We found no differences between the risk groups concerning age or gender. For each increase in the risk profile (low, medium to high) we found an increase in pain-, disability-, catastrophising- and kinesiophobia-scores (see table 2).

The PT treated the patients generally in line with the clinical practice guideline (CPG) on neck pain in applying mobilisation, exercise therapy and giving information.⁸ The GP's do not have a CPG. The GPs referred 22 (out of 26) patients to a PT, one for imaging and 2 to a medical specialist due to persisting pain. In addition, five patients were prescribed pain medication and 15 patients were given some exercises by the GP.

Construct validity We found a very low correlation between activity-question 3 and the NDI-scores ($r = -0.13$). We found moderate correlations between activity-question 4 and the NDI, kinesiophobia- question 5 and the TSK and catastrophizing-questions 6 and 7 and the PCS, and a high correlation between SBT question 1 and the single item question ($r \geq 0.5$) and bothersome-question 9 and the NPRS ($r \geq 0.5$) (table 3). All correlations were as expected a priori regarding the direction of the correlation and the magnitude with the exception of activity-question 3. We conclude that the construct validity is sufficient.

Reproducibility In total, 44 patients completed the second test-retest questionnaire of which 21 patients were regarded stable. On average there were 12 days between T0 and T1.

The quadratic Kappa for the SBT-Neck of 0.58 showed a moderate reproducibility. Distribution is skewed due to the large proportion of patients with "low-risk" and the absence of a stable "high-risk" group. For the "low-risk" group we found a specific agreement of 90.9% and for "medium-risk" 66.7%. We were unable to calculate the specific agreement for the "high-risk" as there were no patients in this 'reproducibility sample'. The overall agreement showed an excellent reproducibility for "low-risk" and a fair reproducibility for the "medium-risk" group.

Predictive validity In total 93 patients completed the T2 questionnaire. In all three risk profiles we found a decrease in pain and disability over time (see table 4). In absolute numbers we found that with the increase of the risk-profile a higher level of pain and disability at three months.

At three months the RR for "medium-risk group" compared to "low-risk group" was 1.5 (95% CI 0.9 – 2.4) for persisting disability, 1.0 (95% CI 0.7 – 1.5) for persisting pain and 0.70 (95% CI 0.5 – 1.1) for perceived recovery. Comparing the "high-risk group" to the low risk group the RR were 1.5 (95% CI 0.5 – 4.1) for persisting disability, 1.2 (95% CI 0.5 – 2.7) for persisting pain and 1.1 (95% CI 0.6 – 2.0) for perceived recovery. An RR of 1.5 means that patients with "medium-risk" had 1.5 times higher chance for persisting neck pain compared to patients with "low-risk". The confidence intervals for all RR include 1 (= equal risks) making it statistically insignificant.

Content validity We analysed data of 100 patients concerning the SBT-Neck in determining floor and ceiling effects. One patient (1.0%) scored zero and one patient (1.0%) scored nine points implying no important floor and ceiling effects and therefore a good content validity.

DISCUSSION

Main findings

The SBT-Neck is a formative model aiming to give a prognosis for future poor disability. The construct validity is sufficient, although activity-question 3 did not meet the a priori expected correlation with the NDI. The test-retest reliability is moderate to almost perfect with a fair kappa. The predictive validity, based on the MCID, showed statistically insignificant results. The absences of floor and ceiling effects confirmed a good content validity.

Interpretation of findings

The specific agreement analysis used to determine the reproducibility, shows highly accurate consistency for patients with a “low-risk” score. The accuracy decreases but is still good for the “medium-risk” group. The conditions were set to ensure ‘stable patients’ based on time, pain and reported recovery. Unfortunately, patients took on average 12 days to respond instead of the aimed 3 days. The delay in response in combination with the range in pain (NPRS +/-1) and recovery (GPE ‘no change’ +/- 1) could result in patients slightly improving and therefore changing to a lower risk-profile. This might explain the lower score in specific agreement for the “medium-risk” group. In interpreting the kappa we have to keep in mind that the distribution is skewed due to the large proportion of patients with “low-risk” and the absence of a “high-risk” group. Nevertheless the SBT-neck is fairly able to distinguish between risk groups.²⁹

We used RR to calculate the increased risk of the medium or high risk groups compared to the “low-risk”-group. All RR calculations using the MCID of the NPRS, the NDI or using the GPE were found to be statistically insignificant. The predictive validity was determined while the clinician applied ‘usual care’. In this cohort no standardized therapy protocol was used as the clinician was blinded for the SBT-score. In contrast to low back pain, no stratified care approach is available for patients with neck pain. We expected the GP or PT to follow the national guidelines when available but during the study only guidelines on trauma-related neck pain were available.^{36,37} In other cases the clinician was free to apply their therapy in the way they deemed fit.

Findings in the context of other literature

This is the first publication on the SBT for patients with neck pain. One study published a version of the SBT that can be used for neck pain but it was specifically designed for multiple body regions.²⁰ Also no stratified care was applied to these patients. Other studies using the SBT focus on low back pain.^{9,12-19} The validation process of this study is comparable to that of the Dutch low back pain version which is also carried out within the PRINS-study, and the results are largely comparable. In the neck pain part of the

PRINS-study there are slightly more people in the “low-risk” group and less in the “high-risk” group compared to the low back pain part.³⁸ This difference is even larger when compared to the initial UK-version on low back pain.⁹

For the construct validity the scores are comparable with the low back pain results except for question 3 (correlation with the NDI) where we found a lower than expected correlation. This might be due to the fact that this is the question we changed from the original version. In the original version the question was “I have only walked short distances because of my back pain”. In the modification we altered this question to “I have used my arms and neck less due to my neck pain” The expert group considered this to be more relevant. This, however, is not a question covered by the NDI and possibly explaining the low correlation.

For the reproducibility we found that the quadratic kappa is slightly lower in this study with 0.58 compared to 0.65 for low back pain, the specific agreement is slightly higher in neck pain.

For the predictive validity we found that both SBT for neck pain and low back pain are better in predicting persisting disability than in predicting persisting pain or recovery. This is in accordance with the aim of the original SBT for low back pain which aimed to predict persisting disability.⁹

The initial SBT study by Hill et al, and also all the translations, used an internal consistency analysis and a psychosocial subscale and calculated a discriminant validity using an Area under the Curve.^{9,12,13,39,40} In our study we approached the tool as a formative model making this calculation redundant.

Strengths and limitations

The strength of this study is that we are the first study that successfully modified the SBT to a neck pain version. The construct validity, reproducibility, and content validity are all moderate to good. The predictive validity is insufficient when using ‘usual care’ as treatment instead of a targeted treatment. The advised minimum sample size was met except for the reproducibility (44 instead of 50). Another limitation of this study is that, for reasons of feasibility, the baseline questionnaire was filled in after the first consultation instead of before as intended. During the initial consultation patient’s cognition might be altered and therefore influence the results.

Clinical and/or research implications

The SBT-Neck needs further research to determine if stratified care can be added and will lead to a faster recovery.

CONCLUSION

The SBT is successfully modified to fit patients with neck pain in Dutch primary care. The psychometric analysis showed sufficient valid and reliable outcomes with the exception

ACKNOWLEDGEMENTS

The authors would like to thank all of the GP's and PT's that included patients for this study. A special thank goes out to Nynke Wildervanck for her contribution and to Steven Constandse, Guido Iken, Joost van Broekhoven and Frans van der Kooij for their extra efforts to reach the needed sample size.

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interests regarding the publication of this paper. This study was undertaken with financial support of CZ healthcare insurance company and Dutch Arthritis Foundation.

REFERENCES

1. Vos T, Flaxman AD, Naghavi M, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*. 2012;380(9859):2163-2196. doi:10.1016/S0140-6736(12)61729-2.
2. Koppes D. Zorg door de fysiotherapeut - top-10 gezondheidsproblemen (DCSPH). www.nivel.nl/node/4677. <http://www.nivel.nl/node/4677>. Published 2016. Accessed November 7, 2016.
3. Hoy DG, Protani M, De R, Buchbinder R. The epidemiology of neck pain. *Best Pr Res Clin Rheumatol*. 2010;24(6):783-792. doi:10.1016/j.berh.2011.01.019.
4. Hush JM, Lin CC, Michaleff ZA, Verhagen A, Refshauge KM. Prognosis of acute idiopathic neck pain is poor: a systematic review and meta-analysis. *Arch Phys Med Rehabil*. 2011;92(5):824-829. doi:10.1016/j.apmr.2010.12.025.
5. Vos CJ, Verhagen AP, Passchier J, Koes BW. Clinical course and prognostic factors in acute neck pain: an inception cohort study in general practice. *Pain Med*. 2008;9(5):572-580. doi:10.1111/j.1526-4637.2008.00456.x.
6. Borghouts JA, Koes BW, Vondeling H, Bouter LM. Cost-of-illness of neck pain in The Netherlands in 1996. *Pain*. 1999;80(3):629-636. <http://www.ncbi.nlm.nih.gov/pubmed/10342424>.
7. Guzman J, Hurwitz EL, Carroll LJ, et al. A new conceptual model of neck pain: linking onset, course, and care: the Bone and Joint Decade 2000-2010 Task Force on Neck Pain and Its Associated Disorders. *Spine (Phila Pa 1976)*. 2008;33(4 Suppl):S14-23. doi:10.1097/BRS.0b013e3181643efb.
8. Bier JD, Scholten-Peeters W, Pool JJ, Staal JB, Tulder MW van, Verhagen AP. *KNGF-Richtlijn Nekpijn*; 2016.
9. Hill JC, Dunn KM, Lewis M, et al. A primary care back pain screening tool: Identifying patient subgroups for initial treatment. *Arthritis Rheum*. 2008;59(5):632-641. doi:10.1002/art.23563.
10. Foster NE, Hill JC, Hay EM. Subgrouping patients with low back pain in primary care: are we getting any better at it? *Man Ther*. 2011;16(1):3-8. doi:10.1016/j.math.2010.05.013.
11. Hill JC, Whitehurst DGT, Lewis M, et al. Comparison of stratified primary care management for low back pain with current best practice (STarT Back): a randomised controlled trial. *Lancet (London, England)*. 2011;378(9802):1560-1571. doi:10.1016/S0140-6736(11)60937-9.
12. Abedi M, Manshadi FD, Khalkhali M, et al. Translation and validation of the Persian version of the STarT Back Screening Tool in patients with nonspecific low back pain. *Man Ther*. 2015:1-5. doi:10.1016/j.math.2015.04.006.
13. Piironen S, Paananen M, Haapea M, et al. Transcultural adaption and psychometric properties of the STarT Back Screening Tool among Finnish low back pain patients. *Eur Spine J*. February 2015. doi:10.1007/s00586-015-3804-6.
14. Pilz B, Vasconcelos RA, Marcondes FB. The Brazilian version of STarT Back Screening Tool – translation , cross-cultural adaptation and reliability *. 2014;18(5):453-461.
15. Luan S, Min Y, Li G, et al. Cross-cultural Adaptation, Reliability, and Validity of the Chinese Version of the STarT Back Screening Tool in Patients With Low Back Pain. *Spine (Phila Pa 1976)*. 2014;39(16):E974-9. doi:10.1097/BRS.0000000000000413.
16. Bruyere O, Demoulin M, Brereton C, et al. Translation validation of a new back pain screening questionnaire (the STarT Back Screening Tool) in French. *Arch Public Heal*. 2012;70(1):12. doi:10.1186/0778-7367-70-12.
17. Morso L, Albert H, Kent P, et al. Translation and discriminative validation of the STarT Back Screening Tool into Danish. *Eur Spine J*. 2011;20(12):2166-2173. doi:10.1007/s00586-011-1911-6.
18. Karstens S, Krug K, Hill JC, et al. Validation of the German version of the STarT-Back Tool (STarT-G): a cohort study with patients from primary care practices. *BMC Musculoskelet Disord*. 2015;16:346. doi:10.1186/s12891-015-0806-9.
19. Bier JD, Ostelo RWJG, Hooff ML van, Wildervanck N, Koes BW, Verhagen AP. Validity and reproducibility of the translated STarT Back Tool in low back pain patients in Dutch primary care. *Pending Publ*. 2016.
20. Butera KA, Lentz TA, Beneciuk JM, George SZ. Preliminary Evaluation of a Modified STarT Back Screening Tool Across Different Musculoskeletal Pain Conditions. *Phys Ther*. February 2016. doi:10.2522/ptj.20150377.
21. Azimi P, Shahzadi S, Azhari S, Montazeri A. A validation study of the Iranian version of STarT Back Screening Tool (SBST) in lumbar central canal stenosis patients. *J Orthop Sci*. 2014;19(2):213-217. doi:10.1007/s00776-013-0506-y.
22. Hill JC, Afolabi EK, Lewis M, et al. 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*. 2016;6(10):e012445. doi:10.1136/bmjopen-2016-012445.
23. Terwee CB, Bot SDM, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007;60(1):34-42. doi:10.1016/j.jclinepi.2006.03.012.
 24. Hjerstad MJ, Fayers PM, Haugen DF, et al. Studies comparing Numerical Rating Scales, Verbal Rating Scales, and Visual Analogue Scales for assessment of pain intensity in adults: a systematic literature review. *J Pain Symptom Manag*. 2011;41(6):1073-1093. doi:10.1016/j.jpainsymman.2010.08.016.
 25. Vernon H, Mior S. The Neck Disability Index: a study of reliability and validity. *J Manip Physiol Ther*. 1991;14(7):409-415. <http://www.ncbi.nlm.nih.gov/pubmed/1834753>.
 26. Vlaeyen JW, Kole-Snijders AM, Boeren RG, van Eek H. Fear of movement/(re)injury in chronic low back pain and its relation to behavioral performance. *Pain*. 1995;62(3):363-372. <http://www.ncbi.nlm.nih.gov/pubmed/8657437>.
 27. Sullivan MJL, Bishop SR, Pivik J. The Pain Catastrophizing Scale: Development and validation. *Psychol Assess*. 1995;7(4):524-532. doi:Doi 10.1037/1040-3590.7.4.524.
 28. Salen BA, Spangfort E V, Nygren AL, Nordemar R. The Disability Rating Index: an instrument for the assessment of disability in clinical settings. *J Clin Epidemiol*. 1994;47(12):1423-1435. <http://www.ncbi.nlm.nih.gov/pubmed/7730851>.
 29. de Vet HCW, Mokkink LB, Terwee CB, Hoekstra OS, Knol DL. Clinicians are right not to like Cohen's kappa. *Bmj-British Med J*. 2013;346(April):f2125. doi:Artn F2125Doi 10.1136/Bmj.F2125.
 30. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed.; 1988. <http://www.amazon.com/Statistical-Analysis-Behavioral-Sciences-Edition/dp/0805802835>. Accessed March 25, 2016.
 31. Kamper SJ, Ostelo RWJG, Knol DL, Maher CG, de Vet HCW, Hancock MJ. Global Perceived Effect scales provided reliable assessments of health transition in people with musculoskeletal disorders, but ratings are strongly influenced by current status. *J Clin Epidemiol*. 2010;63(7):760-766.e1. doi:10.1016/j.jclinepi.2009.09.009.
 32. Sim J, Wright CC. The Kappa Statistic in Reliability Studies: Use, Interpretation, and Sample Size Requirements. *Phys Ther*. 2005;85(3):257-268. <http://ptjournal.apta.org/content/85/3/257.long>. Accessed March 21, 2016.
 33. Jorritsma W, De Vries GE, Dijkstra PU, Geertzen JHB, Reneman MF. Neck Pain and Disability Scale and Neck Disability Index: Validity of Dutch language versions. *Eur Spine J*. 2012;21:93-100. doi:10.1007/s00586-011-1920-5.
 34. Rubinstein SM, Pool JJ, van Tulder MW, Riphagen II, de Vet HC. A systematic review of the diagnostic accuracy of provocative tests of the neck for diagnosing cervical radiculopathy. *Eur Spine J*. 2007;16(3):307-319. doi:10.1007/s00586-006-0225-6.
 35. Pool JJ, Ostelo RW, Hoving JL, Bouter LM, de Vet HC. Minimal clinically important change of the Neck Disability Index and the Numerical Rating Scale for patients with neck pain. *Spine (Phila Pa 1976)*. 2007;32(26):3047-3051. doi:10.1097/BRS.0b013e31815cf75b.
 36. Minderhoud JM, Keuter EJK, Verhagen AP, Valk M, Rosenbrand CJGM, CBO. *Richtlijn Diagnostiek En Behandeling van Mensen Met Whiplash Associated Disorder I / II*. Nederlandse Vereniging voor Neurologie / CBO; 2008.
 37. Scholten-Peeters GG, Bekkering GE, Verhagen AP, et al. Clinical practice guideline for the physiotherapy of patients with whiplash-associated disorders. *Spine (Phila Pa 1976)*. 2002;27(4):412-422. <http://www.ncbi.nlm.nih.gov/pubmed/11840109>.
 38. Bier JD, Ostelo RWJG, Hooff ML van, et al. Validity and reproducibility of the STarT Back Tool (Dutch version) in low back pain patients in primary care. *Unpubl Obs*. 2016.
 39. Bruyère O, Demoulin M, Beaudart C, et al. Validity and reliability of the French version of the STarT Back screening tool for patients with low back pain. *Spine (Phila Pa 1976)*. 2014;39(2):E123-8. doi:10.1097/BRS.0000000000000062.
 40. Aebischer B, Hill JC, Hilfiker R, Karstens S. German Translation and Cross-Cultural Adaptation of the STarT Back Screening Tool. *PLoS One*. 2015;10(7):e0132068. doi:10.1371/journal.pone.0132068.

Table 1, specific agreement*

		Follow-up (T1)		
		Low	Medium	High
Baseline (T0)	Low	15 (A)	2 (B)	0 (C)
	Medium	1 (D)	3 (E)	0 (F)
	High	0 (G)	0 (H)	0 (I)

* "Low-risk" $A/(A+(B+C+D+G)/2) = 15/16.5 = 90.9\%$

"Medium-risk" $E/(E+(B+H+D+F)/2) = 3/4.5 = 66.7\%$

"High-risk" $I/(I+(C+F+G+H)/2) = 0$

Figure 1, patient flow

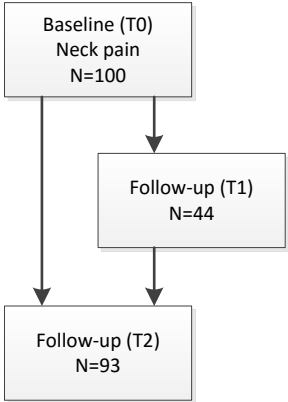


Table 2, baseline characteristics of the study population.*

	Neck pain (N=100)	Low risk (N=58)	Medium risk (N=37)	High risk (N=5)
Female	65 (65.0)	39 (65.5)	23 (62.2)	4 (80.0)
Age in years, mean (SD)	45.6 (14.3)	45.5 (14.1)	44.8 (13.7)	52.8 (21.1)
SBT risk profile				
Low	58 (58.0)			
Medium	37 (37.0)			
High	5 (5.0)			
Episode duration				
<1 month	27 (27.0)	15 (25.9)	11 (29.7)	1 (20.0)
1 to 3 months	18 (18.0)	12 (20.7)	6 (16.2)	0 (0.0)
>3 months	55 (55.0)	31 (53.4)	20 (54.1)	4 (80.0)
SBT score, mean (SD)	3.4 (1.8)	2.1 (0.8)	4.9 (0.9)	7.6 (1.1)
Pain intensity, mean (SD)	5.5 (1.9)	4.7 (1.8)	6.6 (1.4)	7.0 (0.7)
Mild (0-5)	41 (41.0)	34 (58.6)	7 (18.9)	0 (0.0)
Moderate (5-7)	46 (46.0)	24 (41.4)	18 (48.6)	4 (80.0)
Severe (8-10)	13 (13.0)	0 (0.0)	12 (32.4)	1 (20.0)
Disability (NDI), mean (SD)	28.2 (13.6)	22.0 (9.8)	34.6 (12.2)	57.8 (14.2)
Referred pain	33 (33.0)	15 (25.9)	15 (40.5)	3 (60.0)
Comorbid pain	69 (69.0)	36 (62.1)	28 (75.7)	5 (100.0)
Bothersome	48 (48.0)	12 (20.7)	31 (83.8)	5 (100.0)
Fear (TSK), mean (SD)	32.2 (5.9)	29.7 (4.9)	34.8 (4.3)	42.6 (6.9)
Catastrophizing (PCS), mean (SD)	12.7 (10.0)	8.8 (7.3)	16.1 (8.7)	32.6 (15.2)
<p>* Values are numbers (percentage) unless otherwise indicated. pain is measured on the Numeric Pain Rating Scale (0-10). SBT = STarT Back tool, NDI = Neck Disability Index (0-100), TSK = Tampa Scale of Kinesiophobia (17-68), PSC = Pain Catastrophizing Scale (0-65),</p>				

Table 3; Correlation between the STarT Back Tool and their reference questionnaires using the Pearson's correlation

SBT and reference	Correlation			
	A priori	r		Expected
Q1 - sigle item	$r \geq 0.30$	0.55	high	Yes
Q3 - NDI	$r \geq 0.30$	0.13	low	No
Q4 - NDI	$r \geq 0.30$	0.37	moderate	Yes
Q5 - TSK	$r \geq 0.30$	0.42	moderate	Yes
Q6 - PCS	$r \geq 0.30$	0.46	moderate	Yes
Q7 - PCS	$r \geq 0.30$	0.35	moderate	Yes
Q9 - NPRS	$r \geq 0.30$	0.50	high	Yes

SBT = STarT Back Tool, Q = Question, r = Pearson's correlation, NDI = Neck Disability index, TSK = Tampa Scale of Kinesiophobia, PCS = Pain Catastrophizing Scale, NPRS = Numeric Pain Rating Scale

Table 4, relative risk of pain, disability or recovery at three month follow-up *

	Persisting pain			Persisting disability			Recovery		
	%	NPRS	RR (95% CI)	%	NDI	RR (95% CI)	%	N	RR (95% CI)
Low Risk (N=58)	43.6	2.64 (1.94)		65.5	12.32 (9.60)		67.3	37	
Medium Risk (N=37)	55.9	4.26 (2.59)	0.99 (0.68 - 1.45)	50.0	22.00 (15.04)	1.45 (0.88 - 2.38)	47.1	16	0.70 (0.47 - 1.05)
High Risk (N=5)	66.7	5.00 (3.46)	1.18 (0.51 - 2.72)	50.0	35.00 (24.24)	1.45 (0.51 - 4.12)	75.0	3	1.11 (0.61 - 2.02)

* Values are mean scores (SD) unless otherwise indicated. **RR= Relative Risk, CI = confidence interval**, NPRS = Numeric Pain Rating Scale (0-10) , NDI = Neck disability index (0-100)

□

The START Back Screening Tool: Dutch Neck Version

Auteur: D. Bier, B. W. Koes, R. W. J. G. Stelo, B. Mutsaers, N. Wildervanck, A. P. Verhagen (2014)

□

Naam: _____ Datum: _____

Antwoord u is stabiel of niet stabiel in ieder onderdeel. Kruis bij ieder onderdeel het vakje aan dat op u van toepassing is. Soms is het moeilijk om tussen twee vakjes te kiezen, kruis dan het vakje aan dat u het probleem het beste beschrijft. Kruis niet meer dan één vakje per onderdeel aan!

Denk bij het beantwoorden van de volgende vragen telkens aan de situatie in de laatste 2 weken.

□

□	Oneens	Eens
□	□	□
□	□	□

1. In de laatste 2 weken straalde mijn nekpijn wel eens uit naar één of beide armen. □ □ □ □ □ □ □ □ □ □

2. In de laatste 2 weken heb ik, naast mijn nekpijn, wel eens pijn ergens anders gehad. □ □ □ □ □ □ □ □ □ □

3. In de laatste 2 weken bewoog ik mijn nek en/of armen minder vanwege mijn nekpijn. □ □ □ □ □ □ □ □ □ □

4. In de laatste 2 weken kleedde ik me trager dan gewoonlijk aan vanwege mijn nekpijn. □ □ □ □ □ □ □ □ □ □

5. Voor iemand in mijn toestand is het recht niet veilig om lichamelijk actief te zijn. □ □ □ □ □ □ □ □ □ □

6. Ongeruste gedachten gingen vaak door mijn hoofd. □ □ □ □ □ □ □ □ □ □

7. Ik vind dat mijn nekpijn verschrikkelijk is en ik geloof dat het nooit meer beter zal worden. □ □ □ □ □ □ □ □ □ □

8. Over het geheel genomen heb ik niet genoten van alle dingen waar ik vroeger wel van genoot. □ □ □ □ □ □ □ □ □ □

9. Over het geheel genomen, hoe hinderlijk was uw nekpijn in de laatste 2 weken?

□

In het geheel niet	Een beetje	Matig	Erg	Extreem
□	□	□	□	□
□	□	□	□	□

Totale uitslag (alle 9) □ □ □ □ □ □ □ □ □ □ Sub Uitslag (Q5-9): _____

□

□

The Dutch version of the STarT Back Tool for patients with neck pain.

Appendix 2

Original SBT	Dutch neck version	English translation of the Dutch neck version
1 My back pain has spread down my leg(s) at some time in the last 2 weeks	In de laatste 2 weken straalde mijn nekpijn wel eens uit naar één of beide armen.	In the past 2 weeks my neck pain sometimes radiated in one or both arms.
2 I have had pain in the shoulder or neck at some time in the last 2 weeks	In de laatste 2 weken heb ik, naast mijn nekpijn, wel eens pijn ergens anders gehad.	In the past 2 weeks I had pain in other parts of my body next to my neck pain.
3 I have only walked short distances because of my back pain	In de laatste 2 weken bewoog ik mijn nek en/of armen minder vanwege mijn nekpijn.	In the past 2 weeks I moved my neck and / or arm less because of my neck pain.
4 In the last 2 weeks, I have dressed more slowly than usual because of back pain	In de laatste 2 weken kleepte ik me trager dan gewoonlijk aan vanwege mijn nekpijn	In the past 2 weeks I dressed more slowly than usual because of my neck
5 It's not really safe for a person with a condition like mine to be physically active	Voor iemand in mijn toestand is het echt niet veilig om lichamelijk actief te zijn	For someone in my condition it's really not safe to be physically active
6 Worrying thoughts have been going through my mind a lot of the time	Ongeruste gedachten gingen vaak door mijn hoofd.	Worrying thoughts often went through my head.
7 I feel that my back pain is terrible and it's never going to get any better	Ik vind dat mijn nekpijn verschrikkelijk is en ik geloof dat het nooit meer beter zal worden	I think my neck pain is terrible and I believe it will never get better
8 In general I have not enjoyed all the things I used to enjoy	Over het geheel genomen heb ik niet genoten van alle dingen waar ik vroeger wel van genoot	Overall, I have not enjoyed all the things I used to enjoy
9 Overall, how bothersome has your back pain been in the last 2 weeks?	Over het geheel genomen, hoe hinderlijk was uw nekpijn in de laatste 2 weken?	Overall, how bothersome was your neck pain in the past 2 weeks?