PTJ: Physical Therapy & Rehabilitation Journal | *Physical Therapy*, 2021;101:1–8
https://doi.org/10.1093/ptj/pzab211
Advance access publication date September 6, 2021
Original Research

Improving the Predictive Validity of the Dutch STarT Back Tool

Jasper D. Bier, PhD[1,2,*], Milou R. Kuijer, MSc[1,2,3], Annet de Jong, MSc, PT[3], Arianne Verhagen, MSc, PhD[1,4]

[1]Department of General Practice, Erasmus MC, Rotterdam, the Netherlands
[2]FS Fysio, Capelle aan den IJssel, the Netherlands
[3]Hogeschool Utrecht, Utrecht, the Netherlands
[4]Discipline of Physiotherapy, Graduate School of Health, University of Technology Sydney, Sydney, Australia

*Address all correspondence to Mr Bier at: j.bier@erasmusmc.nl

Abstract

Objective. The purpose of this study was to evaluate whether the predictive validity of the Dutch version of the STarT Back Tool (SBT) can be improved by (1) using other cut-off values, (2) changing the items, or (3) adding prognostic factors to the SBT.

Design. This was a secondary analysis of a prospective cohort study (PRINS study: Prevalence of Risk groups in Neck- and back pain patients according to the STarT back screening tool) in patients with low back or neck pain.

Methods. The predictive validity was calculated with a relative risk ratio and a Spearman correlation. The new cut-off values were calculated with receiver operating characteristic curves. Replacing items of the SBT and adding new items were assessed with logistic regression analyses.

Results. A total of 150 patients were included; 51% were categorized as having low risk, 39% as moderate risk, and 11% as high risk. Changing the cut-off total score to $\leq$2 and the subscore to $\geq$5 led to an improvement of the Spearman correlation and RR. Adding the item "duration of the complaints" improved the RR for moderate risk (3.6) (95% CI = 1.6–7.9) and for high risk (9.0) (95% CI = 4.2–19.1) compared with low risk. The new Spearman correlation was improved to $r_s = 0.37$.

Conclusion. The predictive validity was improved by adding the item "duration of the complaints" and changing the cut-off values.

Keywords: Low Back Pain, Physical Therapists, Prognosis

Received: March 4, 2021. **Revised:** May 28, 2021. **Accepted:** August 17, 2021

Introduction

Low back pain (LBP) is a major global health problem.[1,2] LBP can be divided into specific and nonspecific pain; it is called specific when objective pathology exists (eg, fracture, infection, cauda equina, or tumor). In nonspecific LBP, no physically identifiable cause for the pain can be found, which accounts for approximately 90% of the cases.[3]

The costs of LBP in the Netherlands from 2002 until 2007 are estimated to have been between 3.5 and 4.3 billion Euros every year.[4] Most of these costs are caused by people with chronic LBP.[5] A prospective cohort study among the Dutch population (n = 5700) over 10 years showed a prevalence of chronic nonspecific LBP of 20%.[4]

Recovery from nonspecific LBP usually occurs within 3 to 6 weeks. However, for some patients the symptoms become chronic.[3] Older age and gender are important demographic factors related to a poor prognosis.[6,7] Important physical factors for a poor prognosis are the duration and severity of the pain.[6,8,9] Psychosocial factors related to a poor prognosis are pain beliefs, self-efficacy, and psychological or psychosocial stress.[6,9,10] Psychosocial factors often show higher predictive values for persisting LBP than physical factors.[5]

Better targeted treatments ensure better recovery and lower health care costs.[11] To this end, it is essential that general practitioners and physiotherapists are able to screen patients based on (modifiable) prognostic factors for persisting LBP. The Subgroups for Targeted Treatment (STarT) Back Tool (SBT) was developed to provide a better-targeted treatment for LBP patients by categorizing them into 1 of 3 categories: low, moderate, or high risk for poor disability (defined as a score on the Roland Morris Disability Questionnaire [RDQ] of $\geq$7).[12]

The SBT was developed in England (Keele) based on the literature and consensus of a panel of clinicians.[12] It focuses on prognostic factors that can be influenced by treatment. The final 9 items can be grouped into 2 different parts: items 1 to 4 are about physical complaints and items 5 to 9 about psychosocial factors. Physical items are extracted from the RDQ and the psychosocial items from 2 validated questionnaires: the Tampa Scale of Kinesophobia (TSK) and the Pain Catastrophizing Scale (PCS). Besides 4 items about illness beliefs and behavior, there is 1 item concerning depression.[12]

The SBT has been translated into Dutch and shows good reproducibility and construct and content validity.[13] The 3-month predictive validity was based on 150 patients with LBP. The relative risk (RR) was calculated for the moderate-risk group relative to the low-risk group: RR = 1.8 (95% CI = 1.0–3.1) and for the high-risk group compared with the low-risk group: RR = 2.7 (95% CI = 1.4–4.9).[13] Nevertheless, these scores were lower compared with the original predictive validity of the English SBT.[12] This leaves room for improvement, especially with CI of nearly 1.0 for the moderate-risk group, which implicates equal risk relative to the low-risk group.

The cut-off values used were identical to the English version. No study has been found on improving the predictive validity of the SBT by changing the cut-off value, changing SBT items, or adding items to the SBT. Therefore, our aim was to evaluate if the predictive validity of the Dutch SBT (SBT-DV) can be improved by (1) using other cut-off values, (2) changing the items, or (3) adding prognostic factors to the SBT-DV.

Methods

Design

This was a secondary analysis of data of a prospective cohort study (PRINS study: Prevalence of Risk groups in Neck- and back pain patients according to the STarT back screening tool).[13] Details on the design of the study can be found in the initial publication.[13]

Setting

The research was conducted in primary care in the Netherlands. The general practitioners and physical therapists recruited the patients in their practices. Data were collected between November 2014 and May 2015.

Patients

Patients at least 18 years of age who had nonspecific LBP, owned an email address, and were able to speak, write, and understand Dutch were included. Eligible patients were orally informed about the study. After the patients had signed the informed consent, they were registered online by the general practitioner or physical therapist. The patients received an email with a link to the baseline questionnaire.

Intervention

General practitioners and physical therapists were blinded to the results of the patients' questionnaires, including the SBT. Clinicians were asked to treat their patients according to the applicable guidelines. The general practitioner guideline recommends giving advice and, if necessary, simple analgesics in the acute phase of LBP, a time-contingent approach to improving the activity level; if the pain persists, a referral to a physical therapist is recommended.[14]

The physical therapist guideline recommends giving advice and reassurance to patients with a normal course of pain (ie, a decrease in pain in the first 6 weeks). In case of a deviant course of pain, the guideline advises applying an evidence-based intervention (eg, mobilizations or manipulations and exercise therapy). A behavioral treatment by the physical therapist is recommended if the pain is insufficiently reduced after 12 weeks.[15]

Measurements

Baseline

At baseline, the patients filled out a questionnaire with demographic data such as age, gender, and duration of the complaints. Besides the SBT, other questionnaires were used. To assess the average pain during the last week, the Numeric Pain Rating Scale (NPRS) was used, ranging from 0 = "no pain" to 10 = "worst imaginable pain."[16] To assess disability, the RDQ was used, which included 24 statements with the options of 1 = yes or 0 = no. These items result in a sum score that ranges from 0 to 24; a higher score indicates more disability.[17,18] The PCS to assess catastrophizing has 13 statements, each with 5 answer options varying from 0 = "not at all" to 4 = "always." These items result in a sum score that ranges from 0 to 52; a higher score indicates a higher level of catastrophizing.[19] The TSK to assess the fear of movement has 17 statements with a 4-point Likert scale ranging from 1 = "strongly disagree" to 4 = "strongly agree." These items result in a sum score ranging from 17 to 68; a higher score indicates a higher level of kinesiophobia.[20] Finally, the EuroQol-5D (EQ-5D)

was used. The EQ-5D comprises 5 dimensions: mobility, self-care, usual activities, pain and discomfort, and anxiety and depression. The first 5 items have 3 answer options ranging from "no problems" to "severe problems." The sixth item is an overall health status question with an answer range from 0 to 100, from "worst imaginable health" to "best imaginable health."[21]

Follow-Up

Three months after the baseline, the patients received an email with a follow-up questionnaire, including the RDQ and the general perceived effect scale to measure recovery. The general perceived effect is a 7-point Likert scale with answer options ranging from "fully recovered" to "worse than ever."

Data Analysis

All data analyses were performed using SPSS statistics Version 22.

General Analyses

For the original SBT-DV, we calculated the predictive validity. The number and percentages of the patients in each risk profile that were correctly predicted after 3 months were calculated. A high-risk score was well predicted when there was poor disability after 3 months, and a moderate- and low-risk score was well predicted when there was no disability. Poor disability is defined as a RDQ score ≥ 7, which was the median score in the English and Dutch study.[12,13] The RR and Spearman correlation were calculated between the profiles of the SBT (low, moderate, or high risk) and actual persisting disability. A linear correlation between the SBT-DV and poor disability (RDQ ≥ 7) was expected and tested using a multiple logistic regression analysis. The variability of the logistic model was represented by the explained variance (Nagelkerke R^2). To check the SBT-DV for multicollinearity, the correlation between the items of the SBT-DV was calculated. Multicollinearity was present when the correlation was $r > .8$. In that situation, 1 of the 2 items was replaced in the SBT.

Cut-Off Values

Possible new cut-off points of the SBT-DV were determined using receiver operating characteristic curves; the area under the curve (AUC) was calculated. For the first cut-off point, the total score of the SBT-DV was plotted against poor disability to split the low-risk group from the moderate-risk group. For the second cut-off point, the subscore (SBT-DV items 5–9) was plotted against poor disability to split the moderate-risk group from the high-risk group (see Fig. 1). An AUC of 0.5 points to coincidence and an AUC of 1 is perfect. AUC values of 0.70 and higher are considered reasonable, and values of 0.75 and higher are considered good.[22] The first cut-off point will be changed from ≤ 3 to ≤ 2 and ≤ 4, and the second from ≥ 4 to ≥ 3 and 5. The cut-off point with the highest AUC will be taken.

Improving the SBT-DV by Replacing Items

Next, we evaluated whether the SBT-DV could be improved by replacing items. Items from the TSK and PCS within the same construct as the initial item in the SBT were selected (see Tab. 1). First, the items were dichotomized to fit into the model of the SBT. Scores 1 and 2 on the TSK item became "disagree," and scores 3 and 4 became "agree." Scores 0 and 1 of the PCS item became "disagree," and scores 2 to 4 became "agree." Negatively formulated questions were reversed. A Spearman correlation was calculated between the items and persisting disability. Using a logistic regression analysis, we evaluated which item showed the highest correlation. If a new item of the TSK and PCS had a higher odds ratio than the original item in the SBT, it was replaced.

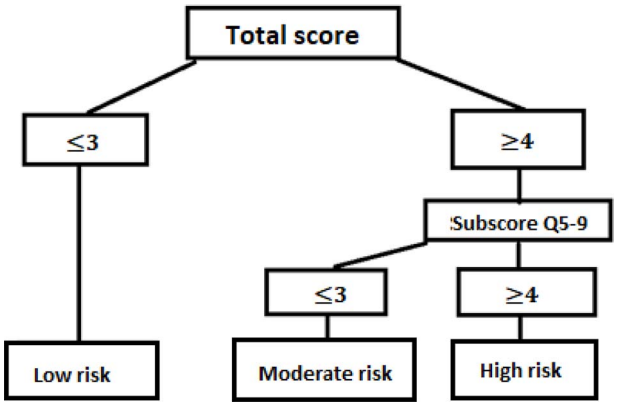

Figure 1. Original cut-off points SBT.

Add New Items

We evaluated whether the SBT could be improved by adding new items. The following possible prognostic factors from the baseline measurement were used: "age" (continuous scale), "gender" (dichotomy scale), "duration of the complaints" (continuous scale in weeks), "degree of pain" (numeric scale 0–10), and "health status" (continuous scale). First, the items were dichotomized to fit the model of the SBT-DV. "Duration of the complaints" was divided into <6 weeks and ≥6 weeks based on a generally accepted division between acute and non-acute.[7] For age (divided into <45 and ≥45 years), degree of pain (divided into <6 and ≥6), and health status (divided into <75 and ≥75 points), we used the median of the original Dutch study.[9] A Spearman correlation was calculated between each item and persisting disability. The item was added to the original model of the SBT, and a new model using logistic regression analysis with persisting disability was calculated. The explained variance of the new model will be represented as Nagelkerke R^2. The new cut-off points were than calculated.

Results

Patient Population

We used the data from 184 patients; another 34 patients did not respond to the follow-up after 3 months and were excluded. Concerning baseline characteristics, the excluded patients did not differ significantly from the individuals who completed the study. At baseline, 76 (51%) participants scored as low risk, 58 (39%) participants as moderate risk, and 16 (11%) participants as high risk on the SBT-DV.

Original Scores

At 3 months, the SBT-DV correctly predicted 79% of the low-risk group, 62% of the moderate-risk group, and 56% of the high-risk group. The Spearman correlation between the 3 risk profiles of the SBT-DV at baseline and persisting disability was $r = .25$.

Table 1. Item SBT and Replaced Item

	Original SBT	Questionnaire, Items[a] (Construct)	Dutch Version of Replaced Item	English Version of Replaced Item
5	It's not really safe for a person with a condition like mine to be physically active	TSK Items 1, 4, 10, 12, 13, 14, 17 (activity avoidance)	TSK Item 17 Ik zou geen lichaamsoefeningen hoeven te doen wanneer ik pijn heb	I should not have to exercise when I am in pain
6	Worrying thoughts have been going through my mind a lot of the time	PCS Items 8, 9, 10, 11 (rumination)	PCS Item 8 Als ik pijn heb verlang ik hevig dat de pijn weggaat	I anxiously want the pain to go away
7	I feel that my back pain is terrible and it's never going to get any better	PCS Items 6, 7, 13 (magnification)		
8	In general, I have not enjoyed all the things I used to enjoy	PCS Items 1, 2, 3, 4, 5, 12 (helplessness)	PCS Item 3 Als ik pijn heb ik dat verschrikkelijk en denk ik dat het nooit beter zal worden	It's terrible, and I think it's never going to get any better

[a]The numbers refer to the items of the relevant questionnaire and are similar in construct as the question of the SBT. PCS = Pain Catastrophizing Scale; SBT = STarT Back Tool; TSK = Tampa Scale for Kinesophobia.

A multivariable logistic regression model shows that the explained variance of the SBT-DV is $R^2 = 0.26$. Multicollinearity between the single items of the SBT was not found. Items 3 through 8 had a significant correlation with persisting disability between $r = .13$ and .30.

Cut-Off Values

The AUC was calculated to evaluate the first cut-off point between the low-risk group and the others. A cut-off point of ≤2 gave an AUC of 0.64, ≤3 gives an AUC of 0.62, and ≤4 gave an AUC of 0.60. The second cut-off point between moderate risk and high risk of ≥3 gave an AUC of 0.54, ≥4 gave an AUC of 0.55, and 5 gave an AUC of 0.56. We took ≤2 for the first cut-off point and 5 for the second. Due to the new cut-off values, the risk distribution changed. The low-risk group reduced to 44 (29%), the number of patients who scored as moderate risk increased to 101 (67%), and 5 (3%) patients scored as high risk at baseline.

With this new distribution, the SBT-DV predicted correctly 89% of the low-risk group, 63% of the moderate-risk group, and 100% of the high-risk group after 3 months. The new RR between low-risk and moderate-risk was calculated again and became RR = 3.3 (95% CI = 1.4–7.7) and between low-risk and high-risk RR = 8.8 (95% CI = 3.9–20.1). The new Spearman correlation was $r = .33$. This means that the predictive validity with the different cut-off points increased.

Improving the SBT-DV by Replacing Items

SBT items were replaced to look for higher correlations between the SBT-DV and poor disability after 3 months. Items of the SBT-DV were replaced when another item with the same construct had a higher correlation with poor disability. Item 5 (construct: activity avoidance) was replaced with item 17 of the TSK. Item 6 (construct: rumination) was replaced for item 8 of the PCS. Item 7 (construct: magnification) had the highest correlation and was not replaced. Item 8 (construct: helplessness) was replaced for item 3 of the PCS (see Tab. 1).

A multivariable logistic regression model with the new items was calculated (see Tab. 2). The explained variance of the new model was $R^2 = 0.29$, higher than the original model.

The predictive validity of the SBT-DV was calculated again. After the items were replaced, 51 (34%) participants scored as low risk, 96 (64%) participants scored as moderate risk, and 3 (2%) participants scored high-risk at baseline. With this new distribution, the SBT-DV correctly predicted 84% of the low-risk group, 63% of the moderate-risk group, and 100% of the high-risk group after 3 months. The new RR for the moderate-risk group decreased to RR = 1.4 (95% CI = 0.8–2.5) and increased for the high-risk group to RR = 2.9 (95% CI = 1.7–5.0). The Spearman correlation was $r = .23$, which is also lower compared with the original model and compared with changing the cut-offs. This means that the predictive validity with the replaced items decreased.

Add New Items

A significant correlation between the new dichotomized items and persisting disability was found for age ($r_s = .2$) and duration of the complaints ($r_s = .3$). A logistic regression with the SBT-DV including the items and persisting disability was calculated. Logistic regression analyses with the SBT-DV and the items gave an improvement of the explained variance: age to $R^2 = 0.31$, duration of the complaints to $R^2 = 0.38$, and age + duration of the complaints to $R^2 = 0.39$. The possible new cut-off points were calculated as ≤3 for the first one and ≥5 for the second if only 1 item was added to the SBT-DV. When both of them are added, the first possible new cut-off point changed to ≤4. Adding the item "duration of the complaints" gave the best predicted values. The new distribution (see Tab. 3) improves the RR for both the moderate-risk group (RR = 3.6 [95% CI = 1.6–7.9]) and the high-risk group (RR = 9.0 [95% CI = 4.2–19.1]) compared with the low-risk group. The new Spearman correlation between the profiles and persisting disability was improved to $r_s = .37$. This means that the predictive validity increased after the item "duration of the complaints" was added and the cut-off values were changed, which is a suggestion for a new version of the SBT-DV (see Figs. 2 and 3).

Discussion

The SBT is a formative model aiming to provide a prognosis on poor disability. The predictive validity improved by changing the first cut-off score between the low-risk group from ≤3 to ≤2 and the second cut-off score between the moderate-risk

Table 2. Multi-Variable Logistic Regression[a]

	Original SBT			Alternative SBT	
	OR	SE		OR	SE
SBT Q1	1.30	0.42		1.41	0.42
SBT Q2	1.52	0.51		1.37	0.49
SBT Q3	2.69[b]	0.46		2.79[b]	0.45
SBT Q4	1.16	0.46		1.22	0.48
SBT Q5	0.98	0.69	TSK 17	2.48	0.52
SBT Q6	1.34	0.42	PCS 8	1.95	0.49
SBT Q7	2.87[b]	0.50		2.17	0.54
SBT Q8	3.00[b]	0.51	PCS 3	3.80[b]	0.61
SBT Q9	0.52	0.53		0.41	0.49
Nagelkerke R^2	0.26			0.29	

[a] OR = odds ratio; PCS = Pain Catastrophizing Scale; SBT = STarT Back Tool; SE = standard error. [b] Significant .05; SBT Q5, Q6 and Q8 are used in model 1; TSK17, PCS 8 and PCS 3 are used in model 2.

Table 3. Adding New Items to the SBT[a]

New Distribution With New Cut-off Points With Adding "Duration of the Complaints"			
Profile SBT	Number of Patients Without Poor Disability After 3 Months	Number of Patients With Poor Disability After 3 Months	% Well Predicted
Low-risk	48	6	89
Moderate-risk	55	36	60
High-risk	0	5	100

[a] SBT = STarT Back Tool.

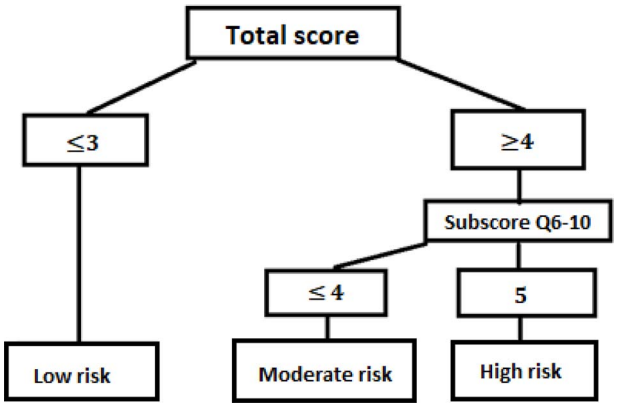

Figure 2. Suggested cut-off points SBT-DV.

group and the high-risk group from ≥4 to 5. The predictive validity of the SBT-DV decreased through the replacement of items but improved through the addition of the items "age" and "duration of the complaints." The suggestion after this study is to improve the SBT-DV by adding the item "duration of the complaints" and changing the first cut-off value to ≤3 and the second to 5.

Interpretation of Findings

The SBT-DV items and cut-off values were taken from the original English version. The items and cut-off values were optimized for the English population, but not for the Dutch, to obtain the best predictive validity. The current research makes it clear that the predictive validity of the Dutch version can be improved by adding a prognostic factor. LBP for more than 6 weeks increases the chance of poor disability after 3 months. The acute phase of LBP lasts 6 weeks; after that, normally the LBP is recovered. The longer the LBP lasts, the greater the chance that psychosocial factors play a role in the poor disability. This corresponds to the high-risk group of the SBT.

The predictive validity also increased by changing the cut-off values. The number of patients in the low-risk group decreased by adding 1 item and leaving the first cut-off point the same. The changing of the second cut-off point also decreased the number of people in the high-risk group and increased the medium-risk group.[23] Both result in better-predicted patients in those groups. *R*-squares are relatively small, but in musculoskeletal research, *R*-squares of 0.26 to 0.39, as we found, are not uncommon.

Findings in the Context of Other Literature

This is the first publication, to our knowledge, in which cut-off values were changed and items replaced in the SBT. The predictive validity of the SBT is published in different languages. The predictive validity of the English version showed a mean RR = 3.2 (95% CI = 2.3–4.4) for moderate risk compared with low risk and RR = 4.7 (95% CI = 3.4–6.4) for high risk compared with low risk. The Spearman correlation between the SBT at baseline and poor disability demonstrated moderate associations ($r = .58$).[12] The predictive validity of the English study remains higher than in this study.[12] The predictive validity of the Danish cohort had an RR = 2.4 (95% CI = 1.7–3.4) for the moderate-risk group and RR = 2.8 (95% CI = 1.8–3.8) for the high-risk group, which is lower than in our study and lower than that of the English cohort.[24] Other studies, such as the French, German, Persian, Japanese, Chinese, and Finnish, are not comparable, because they used the AUC to calculate the predictive validity.[5,25–29] In the English study, the influence of each item is calculated with an AUC.[12] We used univariate regression analyses to calculate the independence contribution.

Suggested New Version of the SBT-DV

Rugscreenings Instrument

Naam: _____ Datum: _____

Antwoord u alstublieft 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 uw 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.**

		Eens	Oneens
		0	**1**
1	Mijn rugklachten duren momenteel minder dan 6 weken.	□	□
2	In de laatste 2 weken **straalde** mijn rugpijn wel eens **uit naar één of beide benen.**	□	□
3	In de laatste 2 weken heb ik wel eens pijn in mijn **schouder** of **nek** gehad.	□	□
4	Vanwege mijn rugpijn **liep** ik alleen **korte afstanden**.	□	□
5	In de laatste 2 weken **kleedde ik me trager** dan gewoonlijk **aan** vanwege mijn rugpijn.	□	□
6	Voor iemand in mijn toestand is het echt niet veilig om lichamelijk actief te zijn.	□	□
7	**Ongeruste gedachten** gingen vaak door mijn hoofd.	□	□
8	Ik vind dat mijn **rugpijn verschrikkelijk** is en ik geloof dat **het nooit meer beter zal worden**.	□	□
9	Over het geheel genomen heb ik **niet genoten** van alle dingen waar ik vroeger wel van genoot.	□	□

10 Over het geheel genomen, hoe hinderlijk was uw rugpijn in de laatste 2 weken?

In het geheel niet	Een beetje	Matig	Erg	Extreem
□	□	□	□	□
0	0	0	1	1

Totale uitslag (alle 10) : _____ Sub Uitslag (Q6-10):_____

Adapted from:

12. 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.

This is a licensed tool (©2007 Keele University) that may not be modified.The copyright (©2007) of the STarT Back Tool and associated materials is owned by Keele University, the development of which was part funded by Arthritis Research UK:

(i) the tool is designed for use by health care practitioners, with appropriate treatment packages for each of the stratified groups; (ii) the tool is not intended to recommend the use of any particular product. For further information please see http://www.keele.ac.uk/sbst/.

No license is required for non-commercial use. If you would like to incorporate the Dutch version of the STarT Back Tool in any way into commercial product materials, please contact Miranda van Hooff for further advice.

Figure 3. Suggested new version of the SBT-DV. Adapted from: 12. 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. This is a licensed tool (©2007 Keele University) that may not be modified. The copyright (©2007) of the STarT Back Tool and associated materials is owned by Keele University, the development of which was part funded by Arthritis Research UK: (1) the tool is designed for use by health care practitioners, with appropriate treatment packages for each of the stratified groups; (2) the tool is not intended to recommend the use of any particular product. For further information please see http://www.keele.ac.uk/sbst/. No license is required for non-commercial use. If you would like to incorporate the Dutch version of the STarT Back Tool in any way into commercial product materials, please contact Miranda van Hooff for further advice.

Strength and Limitations

A strength of the study is that it is the first study, to our knowledge, to have successfully changed the cut-off values and add prognostic factors to improve the predictive validity of the Dutch version. The RR and the correlation with poor disability both improved by adding "duration of the complaints" and changing the cut-off values. A limitation of the study is that we had a relatively small sample size for changing the cut-off scores. Another limitation is that we had to dichotomize the replaced and added items, which led to loss of information and might make these items less useful. Because the format of the items was different, we had to change the answer options to "agree/disagree."

The starting point of constructing the SBT was to use modifiable factors alone.[12] The goal of the SBT is to apply targeted treatment addressing these modifiable factors. We have added duration, which is a non-modifiable factor, improving the predictive validity.

Clinical and/or Research Implications

Adding the item "duration of the complaints" and changing cut-off scores of the SBT-DV improved the predictive validity. Our findings need to be externally validated. Further research is needed to determine whether this new suggestion for the SBT-DV is cost-effective and leads to better-targeted treatment in practice.

Moreover, psychosocial items such as self-efficacy and body awareness can be of prognostic value and might also contribute to the predictive validity of the SBT-DV, which needs to be evaluated.[6,30,31]

The predictive validity of the SBT-DV improved by adding the item "duration of the complaints" and changing the cut-off scores, but not by replacing items. More research is needed to see if psychosocial items such as self-efficacy and body awareness can also improve the predictive validity of the SBT.

Author Contributions

Concept/idea/research design: J.D. Bier, A. de Jong, A. Verhagen
Writing: J.D. Bier, M.R. Kuijer, A. Verhagen
Data collection: J.D. Bier
Data analysis: M.R. Kuijer, A. Verhagen
Project management: J.D. Bier, A. Verhagen
Consultation (including review of manuscript before submitting): A. de Jong

Funding

There are no funders to report for this study.

Disclosures

The authors completed the ICMJE Form for Disclosure of Potential Conflicts of Interest and reported no conflicts of interest.

References

1. Maher C, Underwood M, Buchbinder R. Non-specific low back pain. *Lancet*. 2017;389:736–747.
2. 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;380:2163–2196.
3. van Tulder M, Becker A, Bekkering T, et al. Chapter 3 European guidelines for the management of acute nonspecific low back pain in primary care. *Eur Spine J*. 2006;15:s169–s191.
4. Van Oostrom SH, Monique Verschuren WM, De Vet HCW, Picavet HSJ. Ten year course of low back pain in an adult population-based cohort—the Doetinchem cohort study. *Eur J Pain*. 2011;15:993–998.
5. Pagé I, Abboud J, O Shaughnessy J, Laurencelle L, Descarreaux M. Chronic low back pain clinical outcomes present higher associations with the STarT back screening tool than with physiologic measures: a 12-month cohort study. *BMC Musculoskelet Disord*. 2015;16:201.
6. Hayden JA, Chou R, Hogg-Johnson S, Bombardier C. Systematic reviews of low back pain prognosis had variable methods and results-guidance for future prognosis reviews. *J Clin Epidemiol*. 2009;62:781–796.
7. Swinkels ICS, Wimmers RH, Groenewegen PP, van den Bosch WJH, Dekker J, van den Ende CHM. What factors explain the number of physical therapy treatment sessions in patients referred with low back pain: a multilevel analysis. *BMC Health Serv Res*. 2005;5:74.
8. Grotle M, Foster N, Dunn K, Croft P. Are prognostic indicators poor outcome different for acute and chronic low back pain consulters in primary care? *Pain*. 2010;151:790–797.
9. Campbell P, Foster NE, Thomas E, Dunn KM. Prognostic indicators of low back pain in primary care: five-year prospective study. *J Pain*. 2013;14:873–883.
10. Main CJ, Foster N, Buchbinder R. How important are back pain beliefs and expectations for satisfactory recovery from back pain? *Best Pract Res Clin Rheumatol*. 2010;24:205–217.
11. Suman A, Schaafsma FG, Elders PJ, van Tulder MW, Anema JR. Cost-effectiveness of a multifaceted implementation strategy for the Dutch multidisciplinary guideline for nonspecific low back pain: design of a stepped-wedge cluster randomised controlled trial. *BMC Public Health*. 2015;15:522.
12. Hill JC, Dunn KM, Lewis M, et al. A primary care back pain screening tool: identifying patient subgroups for initial treatment. *Arthritis Rheum*. 2008;59:632–641.
13. Bier JDD, Ostelo RWJGWJG, Koes BWW, Verhagen APP. Validity and reproducibility of the modified STarT Back Tool (Dutch version) for patients with neck pain in primary care. *Musculoskelet Sci Pract*. 2017;31:22–29.
14. NHG-Werkgroep Aspecifieke lagerugpijn. NHG-Standaard Aspecifieke lagerugpijn (tweede herziening). *Huisarts Wet*. 2017;60:78–84.
15. Staal JB, Hendriks EJM, Heijmans M, et al. *Koninklijk Nederlands Genootschap voor Fysiotherapie (KNGF) -richtlijn lage rugpijn*. 07. 2013. Accessed November 8, 2021. http://med-info.nl/Richtlijnen/Bewegingsapparaat/Aspecifieke%20Lage%20Rugklachten.pdf.
16. Hjermstad 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;41:1073–1093.
17. Brouwer S, Kuijer W, Dijkstra PU, Goeken LN, Groothoff JW, Geertzen JH. Reliability and stability of the Roland Morris Disability Questionnaire: intra class correlation and limits of agreement. *Disabil Rehabil*. 2004;26:162–165.
18. Roland M, Morris R. A study of the natural history of low-back pain. Part II: development of guidelines for trials of treatment in primary care. *Spine (Phila Pa 1976)*. 1983;8:145–150.
19. Sullivan MJL, Bishop SR, Pivik J. The pain catastrophizing scale: development and validation. *Psychol Assess*. 1995;7:524–532.
20. 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;62:363–372.
21. Salen BA, Spangfort EV, Nygren AL, Nordemar R. The disability rating index: an instrument for the assessment of disability in clinical settings. *J Clin Epidemiol*. 1994;47:1423–1435.

22. Greiner M, Pfeiffer D, Smith RD. Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. *Prev Vet Med*. 2000;45:23–41.
23. Bier JD,Ostelo RWJG, Koes BW, Verhagen AP. Validity and reproducibility of the modified STarT Back Tool (Dutch version) for patients with neck pain in primary care. *Musculoskelet Sci Pract*. 2017;31:22–29.
24. Morsø L, Kent P, Albert HB, Hill JC, Kongsted A, Manniche C. The predictive and external validity of the STarT back tool in Danish primary care. *Eur Spine J*. 2013;22:1859–1867.
25. 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;20: 1–5.
26. Matsudaira K, Oka H, Kikuchi N, Haga Y, Sawada T, Tanaka S. Psychometric properties of the Japanese version of the STarT back tool in patients with low back pain *PLoS One*. 2016;11:e0152019.
27. 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:E974–E979.
28. 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*. 2015;25: 287–295.
29. Aebischer B, Hill JC, Hilfiker R, Karstens S. German translation and cross-cultural adaptation of the STarT back screening tool. *PLoS One*. 2015;10:e0132068.
30. Foster NE, Delitto A. Embedding psychosocial perspectives within clinical management of low back pain: integration of practice—challenges and opportunities. *Phys Ther*. 2011;91:790–803.
31. Foster NE, Thomas E, Bishop A, Dunn KM, Main CJ. Distinctiveness of psychological obstacles to recovery in low back pain patients in primary care. *Pain*. 2010;148:398–406.