






RESEARCH ARTICLE OPEN ACCESS

The STarT Back Screening Tool: The Nepali Translation, Cross-Cultural Adaptation and Measurement Properties in Adults With Non-Specific Low Back Pain

Riju Maharjan¹  | Sunee Bovonsunthonchai¹  | Roongtiwa Vachalathiti¹  | Ritu Basnet² | Anupa Pathak³  | Jonathan Hill⁴ | Saurab Sharma^{5,6,7} 

¹Faculty of Physical Therapy, Mahidol University, Salaya, Thailand | ²Department of Psychiatry and Mental Health, Maharajgunj Medical Campus, T.U. Institute of Medicine, Maharajganj, Nepal | ³School of Public Health, Faculty of Medicine and Health, University of Sydney, Sydney, Australia | ⁴School of Medicine, Faculty of Medicine and Health Sciences, Keele University, Keele, UK | ⁵Michael J Cousins Pain Management and Research Centre, Royal North Shore Hospital, Sydney, Australia | ⁶School of Health Sciences, Faculty of Medicine and Health, University of New South Wales, Sydney, Australia | ⁷Kolling Institute, Faculty of Medicine and Health, The University of Sydney, Sydney, Australia

Correspondence: Roongtiwa Vachalathiti (roongtiwa.vac@mahidol.edu) | Saurab Sharma (saurab.sharma@health.nsw.gov.au)

Received: 20 August 2024 | **Revised:** 12 September 2024 | **Accepted:** 15 September 2024

Funding: The authors received no specific funding for this work.

Keywords: cross-cultural adaptation | low back pain | measurement properties | musculoskeletal pain | reliability | STarT back screening tool | validity

ABSTRACT

Introduction/Objective: The STarT Back Screening Tool (SBST) stratifies low back pain (LBP) patients based on their risk of chronicity to guide treatment accordingly. The absence of its validated Nepali version limits stratified LBP care in Nepal. The study aimed to translate and cross-culturally adapt the SBST into Nepali and evaluate its measurement properties in adults with LBP.

Methods: The measurement properties of the Nepali SBST were evaluated in 102 Nepali adults with non-specific LBP. We assessed content validity, internal consistency, test-retest reliability, construct and discriminant validity. Item redundancy was evaluated using Cronbach's alpha ($\alpha > 0.90$), test-retest reliability using Intraclass Correlation Coefficient (ICC_{2,1}) and Cohen's kappa using established cutoffs score for categorising patients into risk groups, construct validity using hypothesis testing (if a minimum of 75% of the hypotheses were supported), and discriminant validity using Area Under the Curve (AUC) with the reference scales administered at baseline.

Results: Cronbach's alpha scores were 0.72 for the overall scale and 0.66 for the psychosocial subscale. Test-retest reliability values were good to excellent with ICC_{2,1} of 0.94 (95% CI: 0.87–0.97) for the overall scale and 0.87 (95% CI: 0.73–0.94) for the psychosocial subscale and Kappa values of 0.68 (95% CI: 0.43–0.93) for the overall scale and 0.79 (95% CI: 0.52–1.00) for psychosocial subscale. Construct validity was confirmed as 100% of a priori hypotheses were met. Acceptable discriminative validity was observed with reference scales with AUCs (0.75–0.80).

Conclusions: Nepali SBST demonstrates the reliability and validity of screening for chronicity risk in Nepali adults with LBP. Future studies should evaluate its responsiveness, predictive abilities, and effectiveness in stratifying LBP patients in the Nepalese context.

Abbreviations: AUC, Area Under Curve; BDI, Beck Depression Inventory; CI, Confidence Interval; GRoC, Global Rating of Change; HADS, Hospital Anxiety and Depression Scale; ICC, Intraclass Correlation Coefficient; LBP, Low Back Pain; NICE, National Institute for Health and Care Excellence; NPRS, Numerical Pain Rating Scale; NSLBP, Nonspecific Low Back Pain; ODI, Oswestry Disability Index; PCS, Pain Catastrophizing Scale; PROMIS, Patient-Reported Outcome Measurement Information System; PROMs, Patient-Reported Outcome Measures; RMDQ, Roland Morris Disability Questionnaire; ROC, Receiver Operating Characteristics; SBST, STarT Back Screening Tool; SD, Standard Deviation; STarT, Subgrouping for Targeted Treatment; TSK, Tampa Scale of Kinesiophobia.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2024 The Author(s). *Musculoskeletal Care* published by John Wiley & Sons Ltd.

1 | Introduction

Low back pain (LBP) is a major global health problem, accounting for 619 million prevalent cases and 69 million years lived with disability (YLDs) worldwide (Ferreira et al. 2023). Approximately 84% of LBP cases are classified as non-specific LBP, where the underlying cause remains unknown (Balagué et al. 2012). In Nepal, a lower-middle-income country in South Asia, the prevalence of LBP is estimated to range from 52% to 91%, depending on the population surveyed and the definition of LBP used, and LBP remains the top contributor to disability (Sharma et al. 2019).

Despite the rapid improvement of acute LBP within the first few weeks of onset, persistent low-to-moderate pain and disability affect approximately one-third of people at 1 year (Henschke et al. 2008; Wallwork et al. 2024). Psychological factors, including pain catastrophizing, fear-avoidance beliefs, and depression, significantly contribute to LBP chronicity (Ramond et al. 2011). Chronic LBP leads to an overall poor quality of life affecting overall physical and psychological functioning (Dutmer et al. 2019; Ge et al. 2022). Early identification of individuals at high-risk for poor prognosis can guide targeted LBP treatment strategies.

International LBP clinical practice guidelines recommend stratified approaches for its management using the STarT Back Screening Tool (SBST) (Australian Commission for Safety and Quality in Health Care 2022; UK 2016; Van Wambeke et al. 2017). The SBST identifies patients with treatment-modifiable prognostic indicators (referred leg pain, comorbid pain, disability, bothersomeness, catastrophizing, fear, anxiety, and depression) and classifies them into prognostic categories (low-, medium-, and high-risk groups) for which different clinical decisions can be recommended (Hill et al. 2008). Implementing SBST in primary care for LBP in the United Kingdom has been shown to enhance clinical decision-making, save time and cost, and improve physical and emotional functioning (Foster et al. 2014; Hill et al. 2011; Morsø et al. 2021).

The SBST was originally developed in English and has subsequently been cross-culturally adapted into 41 different languages, but it is not yet available in Nepali. The availability of such a prognostic tool may facilitate a more efficient and effective LBP management in Nepal by streamlining care pathways and ensuring targeted interventions. Therefore, the aim of this study was to translate and cross-culturally adapt SBST into Nepali and to assess its measurement properties in Nepali adults with LBP (see Box 1 for all a priori hypotheses).

2 | Methods

The study was conducted in two stages: (1) linguistic translation and cross-cultural adaptation using standard forward-backward translation guidelines (Beaton et al. 2000), and (2) evaluation of the measurement properties in accordance with COSMIN-based Standards for the Selection of Health Measurement Instruments (COSMIN) (Mokkink et al. 2019; Terwee et al. 2007, 2018). Ethical approvals were obtained from Mahidol

BOX 1 | Summary of hypotheses.

We hypothesised that the Nepali SBST would demonstrate:

1. Content validity.
2. Evidence of no item redundancy (defined as Cronbach's alpha < 0.90) (Abedi et al. 2015; Billis et al. 2021; Hill et al. 2008; Pilz et al. 2014; Raimundo et al. 2017; Wiangkham et al. 2021; Yilmaz Yelvar et al. 2019).
3. Evidence of good-to-excellent test-retest reliability with Intraclass Correlation Coefficient (ICC) and Kappa ≥ 0.70 (Billis et al. 2021; Hill et al. 2008; Luan et al. 2014; Piironen et al. 2016; Raimundo et al. 2017; Wiangkham et al. 2021).
4. Evidence of construct validity via moderate-to-strong positive correlations (Spearman's $\rho \geq 0.30$) with measures of disability, pain catastrophizing, kinesiophobia, and depression and moderate negative correlation with a two-item version of quality of life scale (Billis et al. 2021; Pilz et al. 2017; Wiangkham et al. 2021). Construct validity was deemed achieved if a minimum of 75% of the hypotheses were supported, in accordance with the COSMIN guidelines (Mokkink et al. 2019; Terwee et al. 2007).
5. We hypothesised that the correlations between the psychosocial subscale of the SBST and psychological measures (e.g., Tampa Scale for Kinesiophobia, Pain Catastrophizing Scale, PROMIS Depression Scale-Short Form) would be stronger than the correlations between the total SBST score and the same psychological measures.
6. Evidence of discriminative validity via acceptable to excellent discrimination with an Area Under Curve (AUC) ≥ 0.70 (Hill et al. 2008; Karstens et al. 2015; Pilz et al. 2014).
7. Evidence of no floor and ceiling effects, meaning that participants' scores are not clustered at the lowest or highest possible values.

University (MU-CIRB: 2022/120.0308) and the Nepal Health Research Council (Ref: 3762022). All research processes complied with the Declaration of Helsinki.

2.1 | Stage I: Linguistic Translation and Cross-Cultural Adaptation

A formal 6-step translation process was initiated after obtaining permission to translate the SBST into Nepali from the developer, Professor Hill, a co-author in this paper.

2.1.1 | Forward Translations

Two bilingual native Nepali speakers independently performed forward translations. The first translator is a physiotherapist and an academic at the University of Sydney (AP) with an extensive experience of cross-cultural research in musculoskeletal conditions (Pathak, Sharma, and Jensen 2018; Sharma et al. 2018).

The second bilingual translator (TR) did not have a medical background and was a Red Cross Settlement Advisor in New Zealand during the time of translation.

2.1.2 | Synthesis of Forward Translations

Both forward translators and the principal investigator developed a synthesised version (T3) along with a written report.

2.1.3 | Backward Translations

Two bilingual translators from the United States, without prior knowledge of the English version of the SBST, independently backward translated the synthesised version (T3) from Nepali into English to ensure the accuracy of Nepali translations. One translator was an Information Technology Clinical System Analyst, while the other was an Associate Professor of Physics.

2.1.4 | Expert Committee Meeting

The committee members, consisting of four translators, a linguistic expert, a musculoskeletal physiotherapy expert (RV), two research methodologists (SB and SS), and the lead author (RM) reviewed the original scale and all forward and backward translated versions along with written reports. Discussion continued until a consensus was reached on semantic, idiomatic, experiential, and conceptual equivalences. This resulted in a prefinal version of Nepali SBST. The lead author (RM) maintained close contact with the original developer for reviewing back translations and addressed any issues or concerns raised.

2.1.5 | Cognitive Debriefing

The lead author administered a prefinal version to 15 Nepali adults experiencing non-specific LBP in interview format to assess content validity, which includes comprehensibility, relevance, and comprehensiveness, following COSMIN recommendations (Eremenco, Cella, and Arnold 2005; Terwee et al. 2018). Cultural appropriateness was considered for all items. Seven participants are generally considered adequate for pretesting and cognitive debriefing interviews (Eremenco, Cella, and Arnold 2005; Terwee et al. 2018). The cognitive debriefing sample consisted of adults with non-specific LBP representing diverse age groups, education levels, both sexes, and diverse ethnicities. All the cognitive debriefing interviews were audio-recorded. RM and SS discussed the feedback received during cognitive debriefing and directed questions to JH as needed. All members of the expert committee reviewed and approved the final Nepali version of the SBST (and its final back translation).

2.1.6 | Proofreading

A linguistic expert then reviewed the final Nepali version to check for grammar and sentence structure before finalising the

translated version. Finally, the Nepali SBST, along with all the written reports, was developed and submitted to the original developer.

2.2 | Stage II: Measurement Properties Testing

2.2.1 | Study Design and Settings

This was a longitudinal observational test-retest design. Participants were recruited from the community settings and two tertiary hospitals from September 2022 to December 2022.

2.2.2 | Participants

Participants were Nepali adults with non-specific LBP. Non-specific LBP was defined as pain in the lower back without an identifiable or known specific cause or pathology (Balagué et al. 2012). LBP was further categorised into three groups based on the duration of pain onset, including acute (less than 6 weeks), subacute (between 6 and 12 weeks), and chronic (more than 12 weeks) (Burton et al. 2006). Inclusion criteria were (1) adult participants (18 years or older), (2) fluency in Nepali demonstrated by ability to respond to screening questions, and (3) a clinically meaningful level of pain (Hush et al. 2009) with pain intensity of ≥ 3 on the 11-point Numerical Pain Rating Scale (NPRS) (Sharma et al. 2017). This level of pain intensity allows participants enough room to significantly improve their pain during the course of study. People reporting mild pain of (1 or 2 on NPRS) often consider themselves recovered (Hush et al. 2009). Exclusion criteria were history of spinal surgery, serious spinal pathology (malignancies, infections and fractures), pregnancy, and neurological and rheumatological conditions (e.g., rheumatoid arthritis). A target of 100 participants was determined as adequate for testing measurement properties as per the COSMIN recommendation (Mokkink et al. 2019).

2.2.3 | Measures

2.2.3.1 | The STarT Back Screening Tool (SBST). The Subgrouping for Targeted Treatment, SBST is a validated self-administered tool for screening individuals with LBP for prognostic indicators for persistent, disabling pain. It consists of nine items distributed into two subscales: physical (items 1 to 4: referred leg pain, comorbid pain, and two items related to disability) and psychosocial (items 5 to 9: bothersomeness, catastrophizing, fear, anxiety, and depression) (Hill et al. 2008). All items use a dichotomised response format, except the question on bothersomeness, which is rated on a 5-point Likert scale. Scores range from 0 to 9, with higher scores indicating worse prognosis (Hill et al. 2008). Patients with an overall score ≤ 3 are classified as a 'low-risk group' with three or fewer poor prognostic indicators. Patients scoring ≥ 4 on the overall scale and ≤ 3 on the psychosocial subscale are classified as a 'medium risk group'. The 'high-risk' group includes patients with a score ≥ 4 on the psychosocial subscale, indicating a high risk of persistent disability (Hill et al. 2008). The SBST has good internal consistency (Cronbach's $\alpha = 0.79\text{--}0.86$) (Abedi

et al. 2015; Azimi et al. 2014; Billis et al. 2021; Bruyère et al. 2014; Hasan and Ahmed 2020; Hill et al. 2008; Pilz et al. 2014; Yilmaz Yelvar et al. 2019), test-retest reliability (quadratic weighted Kappa = 0.65–0.79, ICC = 0.73–0.90) (Abedi et al. 2015; Bier et al. 2017; Billis et al. 2021; Bruyère et al. 2014; Hasan and Ahmed 2020; Hill et al. 2008; Karstens et al. 2015; Luan et al. 2014; Maggiani and Abenavoli 2019; Pilz et al. 2014; Raimundo et al. 2017; Robinson and Dagfinrud 2017; Wiangkham et al. 2021; Yilmaz Yelvar et al. 2019), discriminative validity (AUC: Roland Morris Disability Questionnaire (RMDQ) = 0.76–0.92, ODI = 0.56–0.81, PCS = 0.70–0.79, TSK = 0.79–0.80), Hospital Anxiety and Depression Scale (HADS) = 0.66–0.78 (Abedi et al. 2015; Azimi et al. 2014; Billis et al. 2021; Hill et al. 2008; Karstens et al. 2015; Luan et al. 2014; Morsø et al. 2011; Pilz et al. 2017; Wiangkham et al. 2021), and convergence validity (Spearman's ρ : RMDQ = 0.46–0.88, PCS = 0.38–0.41, TSK = 0.30–0.74, HADS = 0.34–0.71 (Abedi et al. 2015; Billis et al. 2021; Bruyère et al. 2014; Hasan and Ahmed 2020; Karstens et al. 2015; Pilz et al. 2014; Wiangkham et al. 2021; Yilmaz Yelvar et al. 2019).

2.2.3.2 | The Numerical Pain Rating Scale (NPRS). The 11-point NPRS is a pain intensity measure where patients rate their pain on a scale of 0–10 (Sharma et al. 2017). The Nepali translation of NPRS has demonstrated excellent test-retest reliability (ICC = 0.80) and good concurrent validity ($r = 0.43$, $p < 0.001$) with Global Rating of Change (GROC) (Sharma et al. 2017). The Nepali NPRS is also sensitive to change with minimum important changes ranging from 1.17 to 1.33 in individuals with musculoskeletal pain (Sharma et al. 2017).

2.2.3.3 | The Global Rating of Change (GROC). The GROC is used to assess the overall self-perceived change in the condition. The Nepali version of GROC has a recommended 7-point model (Kamper, Maher, and Mackay 2009; Sharma et al. 2017). It is a Likert scale with numerical and verbal descriptors, where the midpoint '4' represents 'no change', the left anchor '7' represents 'very much worse' and the right anchor '1' represents 'recovered completely' or 'very much better.' It has demonstrated concurrent validity with pain intensity ($r = 0.43$, $p < 0.001$) (Sharma et al. 2017). Participants reporting 'no change' (GROC = 4) were labelled as stable, and their data were used for the test-retest reliability analysis.

2.2.3.4 | The Oswestry Disability Index Version 2.1b (ODI). The ODI measures functional disability in individuals with LBP. It consists of 10 items, including nine everyday activities of daily living. Each item is rated on a 5-point Likert scale, with higher scores indicating greater disability (Fairbank and Pynsent 2000). The total score is expressed as a percentage of the maximal scores, ranging from 0 to 100. The Nepali translation of ODI has shown good internal consistency (Cronbach's alpha = 0.72) and test-retest reliability (ICC = 0.82) (Acharya et al. 2014).

2.2.3.5 | The Pain Catastrophizing Scale (PCS). PCS measures catastrophic thinking related to pain. It comprises of three subscales: rumination, magnification, and helplessness. Overall, it consists of 13 items rated on a 5-point Likert scale, with higher scores indicating higher levels of pain catastrophizing. The

total score ranged from 0 to 52. The Nepali translation of PCS has demonstrated excellent internal consistency (Cronbach's alpha = 0.90) and test-retest reliability (ICC = 0.92) (Sharma et al. 2018). It has demonstrated concurrent validity with moderate to strong associations with depression and pain intensity ($r = 0.30$ – 0.50) (Sharma et al. 2018).

2.2.3.6 | The Tampa Scale of Kinesiophobia (TSK). The TSK is a widely used tool for assessing fear of pain and movement (Woby et al. 2005). It comprises 17 items rated on a 4-point Likert scale, with higher scores indicating greater fear (Woby et al. 2005). Its Nepali translation has demonstrated good internal consistency (Cronbach's alpha = 0.76) and good test-retest reliability (ICC = 0.82) as well as convergence validity with the RMDQ ($r = 0.50$, $p < 0.001$) (Maharjan et al.; Woby et al. 2005).

2.2.3.7 | The PROMIS Depression Version 1.0 Short Form 8b (PROMIS–Depression). The 8-item Patient-Reported Outcome Measurement Information System (PROMIS) Depression short form assesses depressive symptoms. Respondents rate the frequency of depressive episodes over the past days using a five-point Likert scale. Higher scores indicate a greater frequency of depressive symptoms (HealthMeasures 2011). The Nepali PROMIS Depression short form demonstrated excellent internal consistency (Cronbach's alpha = 0.93), test-retest reliability (ICC = 0.81) and convergence validity with moderate association with pain intensity ($r = 0.38$) (Sharma et al. 2021; Sharma et al. 2018).

2.2.3.8 | Quality of Life (QoL) Rating Scale. The international multidisciplinary panel recommends QoL as a core domain for non-specific LBP (Chiarotto et al. 2018). The QoL rating scale is a 2-item, 5-point Likert scale assessing general QoL and health. Respondents rated their overall quality of life and general health over the past week. The measure has demonstrated acceptable internal consistency (Cronbach's alpha = 0.73) in Nepali (Sharma et al. 2018; Sharma et al. 2019).

2.2.4 | Procedures

Data collection was conducted by the lead investigator (RM) with assistance from a volunteer physiotherapist who received training on study methodology, eligibility screening, obtaining informed consent, questionnaire administration, data management procedures, patient safety, and privacy. Active recruitment of participants was performed through social media, pamphlets, and word of mouth.

Potential participants received an information sheet and consent form, enabling them to ask questions and provide written consent before participation. For those unable to read or sign, verbal consent was obtained with a witness signing on their behalf. In the initial visit, the research assistant and the investigator recorded demographic details and administered the Nepali versions of patient-reported outcomes (SBST, ODI, PCS, TSK, PROMIS Depression v.1.0. Short Form 8b, QoL Rating Scale). These patient-reported outcomes were re-administered 3 days later along with the GROC scale. Participants maintained their ongoing treatment throughout the study duration.

Throughout the process of collecting data, neither the investigator nor the research assistant exerted any influence on the responses provided by the participants.

We checked for missing data to ensure a completed questionnaire. To minimise loss to follow-up, participants were given reminder calls for hospital visits. Those unable to return for follow-up were interviewed over the telephone. We have used these phone calls for follow-up in our previous research (Kc et al. 2019; Kc et al. 2019; Sharma et al. 2020).

2.2.5 | Statistical Analyses

Descriptive statistics were used to illustrate the patients' demographic characteristics. The Kolmogorov–Smirnov test was used to test the normality of the data distribution at a 0.05 level of significance. The following measurement properties were tested in IBM SPSS Statistics (Version 27) (Corp 2017).

2.2.5.1 | Internal Consistency. The internal consistency and item redundancy of the SBST items were evaluated using Cronbach's alpha (α) for the overall scale and psychosocial subscale. We considered Cronbach's α values < 0.70 as poor internal consistency, whereas values > 0.90 as item redundancy. Values ranging from 0.70 to 0.90 were considered as good internal consistency (Streiner, Norman, and Cairney 2024; Terwee et al. 2007).

2.2.5.2 | Test-Retest Reliability. Test-retest reliability was evaluated for participants in the unchanged group (GROC score of 4) by two methods: (1) by calculating the intraclass correlation coefficient based on a single rater, absolute agreement, 2-way random-effects model (ICC_{2,1}) at the 95% CI for the overall scale and psychosocial subscale; and (2) by calculating quadratic weighted Cohen's Kappa for the corresponding risk groups: (i) the overall risk group (low vs. medium/high risk) and (ii) the psychosocial risk group (low/medium vs. high psychosocial risk group) (Hill et al. 2008; Piironen et al. 2016). The following criteria were used to interpret the ICC and Kappa values: poor < 0.40 , fair 0.40–0.59, good 0.60–0.74, and excellent 0.75–1.00 (Cicchetti et al. 2006).

2.2.5.3 | Construct Validity. Construct validity of SBST was evaluated by computing correlations between the SBST (overall scale and psychosocial subscale) and baseline scores of the reference standard questionnaires ODI, PCS, TSK, and PROMIS-Depression Short Form 8b and the QOL rating scale using the Spearman rank correlation coefficient (ρ). The interpretation criteria for correlations were considered strong ($\rho \geq 0.50$), moderate ($0.30 \leq \rho < 0.50$), or weak ($\rho < 0.30$) (Cohen 1988). Construct validity was deemed achieved if a minimum of 75% of the hypotheses were supported, in accordance with the COSMIN guidelines (Terwee et al. 2007).

2.2.5.4 | Discriminative Validity. Discriminative validity was assessed using the Receiver Operating Characteristic (ROC) curve and the area under the curve (AUC). The SBST (overall scale and psychosocial subscale) was compared against reference measures, that is, ODI, PCS, TSK, and PROMIS-Depression

Short Form 8b. Participants were classified as 'cases' experiencing LBP along with disability, catastrophizing, depression, and fear if they scored above specified cutoff scores (ODI ≥ 12 , PCS ≥ 20 , PROMIS-Depression ≥ 60 , TSK ≥ 41) (Hill et al. 2008; Pilz et al. 2017; Tonosu et al. 2012; Van Wyngaarden et al. 2021). Conversely, those who scored below the cutoff scores were considered 'non-cases' indicating LBP without disability, catastrophizing, depression, and fear respectively. AUC values were interpreted as the following: we considered AUC = 0.50 as no discrimination, > 0.50 to < 0.70 as poor discrimination, ≥ 0.70 to < 0.80 as acceptable discrimination, ≥ 0.80 to < 0.90 as excellent discrimination and AUC ≥ 0.90 as outstanding discrimination (Hosmer Jr, Lemeshow, and Sturdivant 2013).

2.2.5.5 | Floor and Ceiling Effects. The presence of floor and ceiling effects was considered if $> 15\%$ of respondents achieved the lowest and highest possible SBST total scores, respectively (Terwee et al. 2007).

3 | Results

3.1 | Linguistic Translation and Cross-Cultural Adaptation (Content Validity)

Fifteen participants with a mean age of 44.5 years (Standard Deviation; SD 15.9) completed the questionnaire during a cognitive debriefing process. On an average, it took 2.1 (SD 0.70) minutes to complete the SBST. The selected participants were from four distinct ethnic groups, which collectively represented the primary ethnic groups in Nepal. The description of the cognitive debriefing sample is presented in Table 1. The participants' mean pain intensity on 0–10 NPRS was 5.73 (SD 1.98).

No major cultural adaptations were needed. Only minor linguistic modifications were made in items 1, 4, 5, 6, and 7. For example '*spread down my legs(s)*' in item 1 was changed to '*moved to my leg*' to achieve semantic equivalence. In item 4 '*dressed more slowly than usual*' was changed into '*wear my clothes*' to match the closet meaning of the item while it captured the upper extremity's physical function achieving semantic and conceptual equivalence. Similarly, in item 7 instead of the word '*terrible*', a translation for the word '*horrible*' was used that is *darlagdo*, which the majority of participants preferred. Online Supporting Information S1 presents the details of the modifications made.

3.2 | Measurement Properties Testing

3.2.1 | Sample Characteristics

The sample consisted of 102 participants: 54 were recruited from the community and 48 were from the hospitals. The mean age of the participants was 40.1 (SD 12.4) years and the mean body mass index was 25.8 (SD 4.2) kg/m². Fifty-nine percent reported chronic LBP, 35% reported acute LBP, and 6% reported subacute LBP. At the follow-up, 22 participants reported slight improvement, 16 reported significant improvement, and 20

TABLE 1 | Description of cognitive debriefing sample.

Variables	Number (%)
Sex	
Women	11 (73.3)
Men	4 (26.7)
Religion	
Hindu	13 (86.7)
Buddhist	2 (13.3)
Ethnicity	
Chhetri	6 (40.0)
Newar	4 (26.7)
Brahmin	2 (13.3)
Others	3 (20.0)
Education	
Higher secondary	5 (33.3)
Undergraduate	6 (40.0)
Master's degree and above	4 (26.7)
Low back pain category	
Acute (< 6 weeks)	6 (40.0)
Subacute (6–12 weeks)	3 (20.0)
Chronic (> 12 weeks)	5 (33.3)
Missing response	1 (6.7)

reported substantial improvement according to the GRoC scale. Thirty-two participants reported no change, while three reported worsening of symptoms based on the GRoC scale. The demographic and clinical characteristics of the study participants are presented in Table 2. The baseline scores of the measurement instruments are presented in Table 3.

3.2.2 | Internal Consistency

Cronbach's α was 0.72 for the overall Nepali SBST scale and 0.66 for the psychosocial subscale.

3.2.3 | Test-Retest Reliability

ICC_{2,1} was 0.94 (95% CI, 0.87–0.97) for an overall scale and 0.87 (95% CI, 0.73–0.94) for the psychosocial subscale when the agreement was calculated using the 32 participants reporting 'no change' in their status on the GRoC scale. Similarly, the corresponding Kappa values for categorised scores were lower: 0.68 (95% CI, 0.43–0.93) for the overall scale and 0.79 (95% CI, 0.52–1.00) for the psychosocial subscale (Table 4).

3.2.4 | Construct Validity

All five hypotheses (100%) were met for both the overall scale and psychosocial subscale of the Nepali SBST, confirming the

construct validity. As hypothesised, the correlations of the SBST psychosocial subscale and the measures of psychological functioning were stronger than the correlations of the SBST total score and the measures of psychological functioning. The strength and direction of associations are presented in Table 5.

3.2.5 | Discriminative Validity

The AUCs for the overall score and psychosocial subscore ranged from 0.73 (95% CI, 0.61–0.85) to 0.80 (0.71–0.89) (see Table 6).

3.2.6 | Floor and Ceiling Effects

No floor or ceiling effects were detected as only 9% of participants reported the lowest scores, and 1% reported the highest scores.

4 | Discussion

The SBST was successfully translated into Nepali and demonstrated content validity, reliability, construct, and discriminative validity without floor and ceiling effects in Nepali adults with non-specific LBP. The results have important clinical, research, and policy implications.

4.1 | Implications

The SBST is the first translation of a LBP screening tool in Nepal. It can be used in clinical settings to identify and manage individuals at high risk of chronic LBP-related disability. By stratifying individuals into risk groups, it can assist clinicians and patients in making informed treatment decisions, potentially leading to improved health outcomes. The availability of Nepali SBST can contribute to future research on targeted interventions to prevent and treat LBP based on risk stratification. The SBST also has the potential to support policymakers in improving healthcare utilization and targeted treatment access.

4.2 | Meaning of the Study Results and Comparisons With Published Research

4.2.1 | Reliability

The Cronbach's alpha values for the overall scale and psychosocial subscale of the Nepali SBST demonstrated no evidence of item redundancy. The SBST is a multi-construct tool developed for screening high-risk individuals with low back pain to stratify them for appropriate treatment, and is not meant to be used as an outcome measure. Therefore, the internal consistency findings for overall and sub-scales should not be evaluated to determine the reliability of the questionnaire. This approach is based on a formative model, where internal consistency holds less significance.

TABLE 2 | Demographic and clinical characteristics of study participants.

Variables	Number (%)
Sex	
Female	65 (63.7)
Male	37 (36.3)
Religion	
Hindu	88 (86.3)
Buddhist	13 (12.7)
Others	1 (1.0)
Ethnicity	
Brahmin	28 (27.5)
Chhetri	24 (23.5)
Newar	24 (23.5)
Tamang/Rai/Limbu/Sherpa/Lama	14 (13.6)
Others	12 (11.8)
Marital status	
Married	75 (73.5)
Unmarried	26 (25.5)
Separated/Divorced	1 (1.0)
Education	
Did not go to school	3 (2.9)
Primary education	5 (4.9)
Secondary	19 (18.6)
Higher secondary	18 (17.6)
Undergraduate	35 (34.3)
Master's degree and above	22 (21.6)
Employment status	
Employed	70 (68.6)
Unemployed/Housemaker	25 (24.5)
Retired/Others	7 (6.9)
Individual monthly income (NPR)	
No income	27 (26.5)
< 10,000	4 (3.9)
11,000–30,000	26 (25.5)
31,000–49,000	25 (24.5)
≥ 50,000	20 (19.6)
Comorbidities	
No comorbidities	69 (67.6)
Hypertension	13 (12.7)
Multiple/Others	13 (12.7)
Hyperthyroidism	5 (4.9)
Diabetes	2 (2.0)
STarT back risk categorisation	
Low risk	53 (52.0)
Medium risk	29 (28.4)

(Continues)

TABLE 2 | (Continued)

Variables	Number (%)
High risk	20 (19.6)
Disability categorisation	
ODI ≥ 12	80 (78.4)
ODI < 12	22 (21.6)
Catastrophizing categorisation	
PCS ≥ 20	25 (24.5)
PCS < 20	77 (75.5)
Depression categorisation	
PROMIS-depression ≥ 60	10 (10)
PROMIS-depression < 60	92 (90)
Fear avoidance categorisation	
TSK ≥ 41	29 (28.5)
TSK < 41	73 (71.5)

Abbreviations: ODI Oswestry Disability Index, PCS Pain Catastrophizing Scale, PROMIS Patient Reported Outcome Measurement Information System, TSK Tampa Scale of Kinesiophobia.

TABLE 3 | Baseline scores of the measurement instruments (N = 102).

Measurement instruments (score range)	Median (IQR) or mean (SD)	Range (min–max)
NPRS score (0–10)	5 (4–8)	3–10
SBST overall score (0–9)	3 (1–6)	0–9
SBST psychosocial subscore (0–5)	2 (1–3)	0–5
ODI score (0–100)	22 (14–36)	2–72
PCS score (0–52)	13 (5.0–19.2)	0–42
TSK score ^a (17–68)	36.9 (0.6)	21–51
PROMIS-depression short form 8b score (37.1–81.1)	47.2 (54.3–37.1)	31.1–73.6
QoL score (2–10)	7 (6–8)	3–10

Abbreviations: IQR Interquartile range, N Sample, NPRS Numerical Pain Rating Scale, ODI Oswestry Disability Index, PCS Pain Catastrophizing Scale, PROMIS Patient Reported Outcome Measurement Information System, QoL Quality of Life, SBST STarT Back Screening Tool, TSK Tampa Scale of Kinesiophobia.

^aTSK scores were normally distributed; therefore, the presented scores are mean and SD.

TABLE 4 | Reliability of the Nepali STarT back screening tool.

Measurement instruments	Cronbach's α (N = 102)	ICC _{2,1} (95% CI) (N = 32)	Kappa (95% CI) (N = 32)
SBST overall scale	0.72	0.94 (0.87–0.97)	0.68 (0.43–0.93)
SBST psychosocial subscale	0.66	0.87 (0.73–0.94)	0.79 (0.52–1.00)

Abbreviations: CI Confidence interval, ICC Intraclass correlation coefficient, N sample, SBST STarT Back Screening Tool.

TABLE 5 | Correlations of the Nepali STarT Back Screening Tool with criterion measures for construct validity ($N = 102$).

Measurement instruments	SBST overall scale		SBST psychosocial subscale	
	ρ (95% CI)	p value	ρ (95% CI)	p value
ODI	0.65 (0.52–0.75)	< 0.001 ^a	0.53 (0.37–0.66)	< 0.001 ^a
PCS	0.48 (0.31–0.62)	< 0.001 ^a	0.55 (0.39–0.67)	< 0.001 ^a
TSK	0.43 (0.25–0.58)	< 0.001 ^a	0.49 (0.32–0.63)	< 0.001 ^a
PROMIS- depression	0.32 (0.13–0.49)	0.001 ^a	0.37 (0.18–0.53)	< 0.001 ^a
QOL	−0.30 (−0.47 to −0.11)	0.002 ^a	−0.37 (−0.53 to −0.18)	0.002 ^a

Abbreviations: ρ Spearman correlation coefficient (ρ), CI Confidence Interval, ODI Oswestry Disability Index, PCS Pain Catastrophizing Scale, PROMIS Patient Reported Outcome Measurement Information System, QOL Quality of Life N Sample, SBST STarT Back Screening Tool, TSK Tampa Scale of Kinesiophobia.

^aStatistically significant correlations ($p < 0.05$).

TABLE 6 | Discriminative validity of the Nepali STarT Back Screening Tool ($N = 102$).

Case definition	AUCs of SBST overall scale		AUCs of SBST psychosocial subscale	
	(95% CI)	p value	(95% CI)	p value
ODI ≥ 12	0.80 (0.71–0.89)	< 0.001 ^a	0.77 (0.66–0.87)	< 0.001 ^a
PCS ≥ 20	0.73 (0.61–0.85)	0.001 ^a	0.77 (0.66–0.87)	< 0.001 ^a
TSK ≥ 41	0.75 (0.63–0.87)	< 0.001 ^a	0.79 (0.69–0.89)	< 0.001 ^a
PROMIS-depression ≥ 60	0.80 (0.67–0.93)	0.002 ^a	0.78 (0.66–0.91)	0.003 ^a

Abbreviations: AUCs Area under the curve, CI Confidence interval, N Sample, ODI Oswestry Disability Index, PCS Pain Catastrophizing Scale, PROMIS Patient Reported Outcome Measurement Information System, SBST STarT Back Screening Tool, TSK Tampa Scale of Kinesiophobia.

^aStatistically significant correlations ($p < 0.05$).

The test-retest reliability was good to excellent for the overall scale and psychosocial subscale. The findings were comparable to the Norwegian (Robinson and Dagfinrud 2017), German (Karstens et al. 2015), Persian (Abedi et al. 2015), and Dutch (Bier et al. 2017) versions of the scale. Our study opted for a 3-day test-retest interval for pragmatic reasons (loss to follow-up is much higher for longer follow-ups in our settings), which we (Basnet et al. 2023) and others have used in previous research (Bier et al. 2017; Luan et al. 2014; Piironen et al. 2016; Robinson and Dagfinrud 2017; Wiangkham et al. 2021). Three-day follow-up also minimises symptom alterations and is also sufficient for preventing recall bias, especially with eight scales administered during baseline and follow-up assessments (Acharya et al. 2014; Sharma et al. 2017; Sharma et al. 2018; Sharma et al. 2018).

4.2.2 | Validity

The Nepali SBST demonstrated construct validity as all a priori hypotheses were supported, consistent with Thai (Wiangkham et al. 2021), Brazilian (Pilz et al. 2017), Greek (Billis et al. 2021), Turkish (Yilmaz Yelvar et al. 2019), Bangla (Hasan and Ahmed 2020), French (Bruyère et al. 2014), and German (Karstens et al. 2015) versions. The Nepali SBST demonstrated discriminative validity as all observed AUCs were acceptable to excellent, consistent with Brazilian (Pilz et al. 2017), English (Hill et al. 2008), German (Karstens et al. 2015), Thai (Wiangkham et al. 2021), and Greek (Billis et al. 2021) versions. These findings suggest that the Nepali SBST effectively differentiated the clinical status of individuals, as measured by some of the commonly used self-reported measures. In general, the observed AUCs in our study were slightly lower than those

previously reported for the original English version of the SBST (Hill et al. 2008). Differences in AUCs across studies may be attributed to variations in reference standard measures used across studies (Billis et al. 2021; Karstens et al. 2015; Wiangkham et al. 2021).

4.3 | Strengths and Limitations

This study has notable strengths. The translation and cross-cultural adaptation processes followed international standards, ensuring that the Nepali SBST is comprehensible, relevant, and culturally appropriate for Nepali-speaking adults. We adhered to COSMIN recommendations, ensuring methodological quality and accurate interpretation of measurement properties (Mokkink et al. 2019; Terwee et al. 2018). For example, adequate sample size and appropriate statistical methods were employed for the assessment of reliability and validity. Test-retest reliability was assessed under the same conditions, using GROc, ensuring participant stability. Construct and discriminative validity were evaluated by testing a priori hypotheses. Additionally, participants were recruited from clinical and community settings, which broadens the generalisability of our findings.

The results of the study should also be considered in the light of its limitations. First, the study included a small proportion of individuals with low education levels and a small number of high-risk participants, limiting the generalisability of the study results to these populations. The limited number of high-risk participants posed difficulty in testing the psychosocial subscale. Second, the study utilised a short follow-up period and

the participants were not followed up at a longer follow-up time-point to test the predictive ability of the SBST. Therefore, future research utilising a longer follow-up (e.g., 6 months) to assess its predictive capability. Finally, whilst we followed guidance from Beaton and colleagues (Beaton et al. 2000), our expert committee did not include people with lived experience of low back pain, which in hindsight, was a clear study limitation.

4.4 | Recommendations for Future Research

Future research should focus on investigating additional measurement properties of the Nepali SBST, such as predictive validity and responsiveness, using longer follow-up time periods. Validating the SBST in other types of LBP beyond non-specific LBP would be beneficial. Given the prevalence of internet and smartphone usage, there is a need to assess the measurement equivalence of the online version of the SBST for the Nepali population. Conducting high-quality randomized trials testing the effectiveness of the stratified approach in lower middle-income country settings such as Nepal may help appropriate resource allocation for the different risk groups and prevent the provision of providing unnecessary care for low/medium risk individuals with LBP (Buchbinder et al. 2018; Sharma and McAuley 2022).

5 | Conclusion

The SBST was successfully translated and cross-culturally adapted into Nepali. The Nepali translation of the SBST demonstrated test-retest reliability, construct validity, and discriminative validity. Items once considered part of the psychosocial subscale have been discovered to represent more than one construct. Therefore, we suggest that this subscale is better perceived as a grouping of individual psychosocial constructs rather than using it as an outcome measure. The Nepali translation of the SBST can be used for research and clinical purposes among Nepali-speaking adults with non-specific LBP. The effectiveness of the stratified approach for LBP management using SBST needs further testing in Nepal.

Author Contributions

R.M., S.S., R.V., S.B., A.P. and J.H. conceived the study. R.M. conducted data collection, data analysis, and interpretations. R.B. guided the pre-testing process. R.M. wrote the manuscript. R.V., S.B. and S.S. supervised all aspects of the study. All co-authors made a significant intellectual contribution to the manuscript. All authors have read and approved the final manuscript.

Acknowledgements

The authors of the study would like to acknowledge (1) Mahidol University for supporting the study; (2) volunteer translators of the STarT Back (Tika Regmi, Jeniya Maharjan, Amit Dangol, and Balram Adhikari); (3) Sofiya Thapa for assisting in data collection; (4) all other physiotherapy colleagues of the HAMS Hospital; and (5) all the study participants who volunteered to participate in the study.

Ethics Statement

The study protocol was approved by the Institutional Review Board of Mahidol University (MU-CIRB 2022/120.0308) and the Nepal Health Research Council (Ref: 3762022). The study complied with the Declaration of Helsinki. All participants provided written informed consent prior to participation.

Consent

The authors have nothing to report.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

- Abedi, M., F. D. Manshadi, M. Khalkhali, et al. 2015. "Translation and Validation of the Persian Version of the STarT Back Screening Tool in Patients With Nonspecific Low Back Pain." *Manual Therapy* 20, no. 6: 850–854. <https://doi.org/10.1016/j.math.2015.04.006>.
- Acharya, R. S., S. Al-Oraibi, S. P. Adhikari, N. Parajuli, H. Limbu, and F. Al Enezi. 2014. "Validation in the Cross-Cultural Adaptation of the Nepali Version of the Oswestry Disability Index." *Indian Journal of Physiotherapy & Occupational Therapy* 8, no. 2: 158. <https://doi.org/10.5958/j.0973-5674.8.2.079>.
- Australian Commission for Safety and Quality in Health Care. 2022. *Low Back Pain Clinical Care Standard*. Sydney: ACSQHC. <https://www.safetyandquality.gov.au/standards/clinical-care-standards/low-back-pain-clinical-care-standard>.
- Azimi, P., S. Shahzadi, S. Azhari, and A. Montazeri. 2014. "A Validation Study of the Iranian Version of STarT Back Screening Tool (SBST) in Lumbar Central Canal Stenosis Patients." *Journal of Orthopaedic Science* 19, no. 2: 213–217. <https://doi.org/10.1007/s00776-013-0506-y>.
- Balagué, F., A. F. Mannion, F. Pellisé, and C. Cedraschi. 2012. "Non-Specific Low Back Pain." *Lancet* 379, no. 9814: 482–491. [https://doi.org/10.1016/s0140-6736\(11\)60610-7](https://doi.org/10.1016/s0140-6736(11)60610-7).
- Basnet, R., M. P. Jensen, A. Pathak, et al. 2023. "Self-Efficacy in Nepali Adults With Musculoskeletal Pain: Measurement Properties of Hard-Copy and Online Versions of the Pain Self-Efficacy Questionnaire." *Journal of Pain* 25, no. 4: 918–933. <https://doi.org/10.1016/j.jpain.2023.10.012>.
- Beaton, D. E., C. Bombardier, F. Guillemin, and M. B. Ferraz. 2000. "Guidelines for the Process of Cross-Cultural Adaptation of Self-Report Measures." *Spine* 25, no. 24: 3186–3191. <https://doi.org/10.1097/00007632-200012150-00014>.
- Bier, J. D., R. W. J. G. Ostelo, M. L. van Hooff, B. W. Koes, and A. P. Verhagen. 2017. "Validity and Reproducibility of the STarT Back Tool (Dutch Version) in Patients With Low Back Pain in Primary Care Settings." *Physical Therapy* 97, no. 5: 561–570. <https://doi.org/10.1093/ptj/pzx023>.
- Billis, E., K. Fousekis, M. Tsekoura, et al. 2021. "Cross-Cultural Validation of the STarT Back Screening Tool in a Greek Low Back Pain Sample." *Musculoskeletal Science and Practice* 53: 102352. <https://doi.org/10.1016/j.msksp.2021.102352>.
- Bruyère, O., M. Demoulin, C. Beaudart, et al. 2014. "Validity and Reliability of the French Version of the STarT Back Screening Tool for Patients With Low Back Pain." *Spine* 39, no. 2: E123–E128. <https://doi.org/10.1097/brs.000000000000062>.

- Buchbinder, R., M. van Tulder, B. Öberg, et al. 2018. "Low Back Pain: A Call for Action." *Lancet* 391, no. 10137: 2384–2388. [https://doi.org/10.1016/S0140-6736\(18\)30488-4](https://doi.org/10.1016/S0140-6736(18)30488-4).
- Burton, A. K., F. Balagué, G. Cardon, et al. 2006. "Chapter 2. European Guidelines for Prevention in Low Back Pain: November 2004." *European Spine Journal* 15, no. Supplement 2: S136–S168. <https://doi.org/10.1007/s00586-006-1070-3>.
- Chiarotto, A., M. Boers, R. A. Deyo, et al. 2018. "Core Outcome Measurement Instruments for Clinical Trials in Nonspecific Low Back Pain." *Pain* 159, no. 3: 481–495. <https://doi.org/10.1097/j.pain.0000000000001117>.
- Cicchetti, D., R. Bronen, S. Spencer, et al. 2006. "Rating Scales, Scales of Measurement, Issues of Reliability: Resolving Some Critical Issues for Clinicians and Researchers." *Journal of Nervous and Mental Disease* 194, no. 8: 557–564. <https://doi.org/10.1097/01.nmd.0000230392.83607.c5>.
- Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale: Lawrence Erlbaum Associates.
- Corp, I. 2017. *IBM SPSS Statistics for Windows*. Version 27.0. New York, Armonk, NY. <https://www.ibm.com/spss>.
- Dutmer, A. L., H. R. S. Preuper, R. Soer, et al. 2019. "Personal and Societal Impact of Low Back Pain: The Groningen Spine Cohort." *Spine* 44, no. 24: E1443–E1451. <https://doi.org/10.1097/BRS.00000000000003174>.
- Eremenco, S. L., D. Cella, and B. J. Arnold. 2005. "A Comprehensive Method for the Translation and Cross-Cultural Validation of Health Status Questionnaires." *Evaluation and Health Professions* 28, no. 2: 212–232. <https://doi.org/10.1177/0163278705275342>.
- Fairbank, J. C. T., and P. B. Pynsent. 2000. "The Oswestry Disability Index." *Spine* 25, no. 22: 2940–2952. <https://doi.org/10.1097/00007632-200011150-00017>.
- Ferreira, M. L., K. de Luca, L. M. Haile, et al. 2023. "Global, Regional, and National Burden of Low Back Pain, 1990–2020, its Attributable Risk Factors, and Projections to 2050: A Systematic Analysis of the Global Burden of Disease Study 2021." *Lancet Rheumatology* 5, no. 6: e316–e329. [https://doi.org/10.1016/S2665-9913\(23\)00098-X](https://doi.org/10.1016/S2665-9913(23)00098-X).
- Foster, N. E., R. Mullis, J. C. Hill, et al. 2014. "Effect of Stratified Care for Low Back Pain in Family Practice (IMPACT Back): A Prospective Population-Based Sequential Comparison." *Annals of Family Medicine* 12, no. 2: 102–111. <https://doi.org/10.1370/afm.1625>.
- Ge, L., M. J. Pereira, C. W. Yap, and B. H. Heng. 2022. "Chronic Low Back Pain and its Impact on Physical Function, Mental Health, and Health-Related Quality of Life: A Cross-Sectional Study in Singapore." *Scientific Reports* 12, no. 1: 20040. <https://doi.org/10.1038/s41598-022-24703-7>.
- Hasan, M. I., and S. M. Ahmed. 2020. "Translation, Cross-Cultural Adaptation and Validation of English Start Back Screening Tool into Bangla for Patients With Low Back Pain." *Bangabandhu Sheikh Mujib Medical University Journal* 13, no. 3: 73–78. <https://doi.org/10.3329/bssmuj.v13i3.49265>.
- HealthMeasures. 2011. *PROMIS Scoring Manuals*. Northwestern University, HealthMeasures. <http://www.healthmeasures.net>.
- Henschke, N., C. G. Maher, K. M. Refshauge, et al. 2008. "Prognosis in Patients With Recent Onset Low Back Pain in Australian Primary Care: Inception Cohort Study." *BMJ* 337, no. 7662: a171. <https://doi.org/10.1136/bmj.a171>.
- Hill, J. C., K. M. Dunn, M. Lewis, et al. 2008. "A Primary Care Back Pain Screening Tool: Identifying Patient Subgroups for Initial Treatment." *Arthritis & Rheumatism* 59, no. 5: 632–641. <https://doi.org/10.1002/art.23563>.
- Hill, J. C., D. G. Whitehurst, M. Lewis, et al. 2011. "Comparison of Stratified Primary Care Management for Low Back Pain With Current Best Practice (STarT Back): A Randomised Controlled Trial." *Lancet* 378, no. 9802: 1560–1571. [https://doi.org/10.1016/S0140-6736\(11\)60937-9](https://doi.org/10.1016/S0140-6736(11)60937-9).
- Hosmer Jr, D. W., S. Lemeshow, and R. X. Sturdivant. 2013. *Applied Logistic Regression*. 3rd ed., Vol. 398. Hoboken, New Jersey: John Wiley & Sons.
- Hush, J. M., K. Refshauge, G. Sullivan, L. De Souza, C. G. Maher, and J. H. McAuley. 2009. "Recovery: What Does This Mean to Patients With Low Back Pain?" *Arthritis & Rheumatism* 61, no. 1: 124–131. <https://doi.org/10.1002/art.24162>.
- Kamper, S. J., C. G. Maher, and G. Mackay. 2009. "Global Rating of Change Scales: A Review of Strengths and Weaknesses and Considerations for Design." *Journal of Manual & Manipulative Therapy* 17, no. 3: 163–170. <https://doi.org/10.1179/jmt.2009.17.3.163>.
- Karstens, S., K. Krug, J. C. Hill, et al. 2015. "Validation of the German Version of the STarT-Back Tool (STarT-G): A Cohort Study With Patients From Primary Care Practices." *BMC Musculoskeletal Disorders* 16, no. 1: 346. <https://doi.org/10.1186/s12891-015-0806-9>.
- Kc, S., S. Sharma, K. Ginn, T. Almadi, and D. Reed. 2019a. "Nepali Translation, Cross-Cultural Adaptation and Measurement Properties of the Shoulder Pain and Disability Index (SPADI)." *Journal of Orthopaedic Surgery and Research* 14, no. 1: 1–10. <https://doi.org/10.1186/s13018-019-1285-8>.
- Kc, S., S. Sharma, K. Ginn, T. Almadi, H. Subedi, and D. Reed. 2019b. "Cross-Cultural Adaptation and Measurement Properties of the Nepali Version of the DASH (Disability of Arm, Shoulder and Hand) in Patients With Shoulder Pain." *Health and Quality of Life Outcomes* 17, no. 1: 1–10. <https://doi.org/10.1186/s12955-019-1105-1>.
- Luan, S., Y. Min, G. Li, et al. 2014. "Cross-Cultural Adaptation, Reliability, and Validity of the Chinese Version of the STarT Back Screening Tool in Patients With Low Back Pain." *Spine* 39, no. 16: E974–E979. <https://doi.org/10.1097/brs.0000000000000413>.
- Maggiani, A., and A. Abenavoli. 2019. "Italian Translation and Cross-Cultural Adaptation of a Back Pain Screening Questionnaire (STarT Back Screening Tool)." *Annali di Igiene: Medicina Preventiva e di Comunità* 31, no. 1: 69–75. <https://doi.org/10.7416/ai.2019.2260>.
- Maharjan, S., M. Pokharel, C. B. Khatri, et al. "Nepali Translation of the Tampa Scale of Kinesiophobia (TSK) Is Reliable, Valid, and Responsive in Adults With Low Back Pain." (In writing).
- Mokkink, L. B., C. A. Prinsen, D. L. Patrick, et al. 2019. *COSMIN Study Design Checklist for Patient-Reported Outcome Measurement Instruments*. Amsterdam, the Netherlands: Consensus-based Standards for the selection of health Measurement Instruments.
- Morsø, L., H. Albert, P. Kent, C. Manniche, and J. Hill. 2011. "Translation and Discriminative Validation of the STarT Back Screening Tool into Danish." *European Spine Journal* 20, no. 12: 2166–2173. <https://doi.org/10.1007/s00586-011-1911-6>.
- Morsø, L., K. Olsen Rose, B. Schiøttz-Christensen, G. Sowden, J. Søndergaard, and D. H. Christiansen. 2021. "Effectiveness of Stratified Treatment for Back Pain in Danish Primary Care: A Randomized Controlled Trial." *European Journal of Pain* 25, no. 9: 2020–2038. <https://doi.org/10.1002/ejp.1818>.
- Pathak, A., S. Sharma, and M. P. Jensen. 2018. "The Utility and Validity of Pain Intensity Rating Scales for Use in Developing Countries." *Pain Reports* 3, no. 5: e672. <https://doi.org/10.1097/pr.0000000000000672>.
- Piironen, S., M. Paananen, M. Haapea, et al. 2016. "Transcultural Adaptation and Psychometric Properties of the STarT Back Screening Tool Among Finnish Low Back Pain Patients." *European Spine Journal* 25, no. 1: 287–295. <https://doi.org/10.1007/s00586-015-3804-6>.
- Pilz, B., R. A. Vasconcelos, F. B. Marcondes, S. S. Lodovichi, W. Mello, and D. B. Grossi. 2014. "The Brazilian Version of STarT Back Screening Tool—Translation, Cross-Cultural Adaptation and Reliability." *Brazilian Journal of Physical Therapy* 18, no. 5: 453–461. <https://doi.org/10.1590/bjpt-rbf.2014.0028>.
- Pilz, B., R. A. Vasconcelos, P. P. Teixeira, et al. 2017. "Construct and Discriminant Validity of STarT Back Screening Tool—Brazilian

- Version." *Brazilian Journal of Physical Therapy* 21, no. 1: 69–73. <https://doi.org/10.1016/j.bjpt.2016.12.006>.
- Raimundo, A., J. Parraça, N. Batalha, et al. 2017. "Portuguese Translation, Cross-Cultural Adaptation and Reliability of the Questionnaire «Start Back Screening Tool» (SBST)." *Acta Reumatologica Portuguesa* 42, no. 1: 38–46.
- Ramond, A., C. Bouton, I. Richard, et al. 2011. "Psychosocial Risk Factors for Chronic Low Back Pain in Primary Care—a Systematic Review." *Family Practice* 28, no. 1: 12–21. <https://doi.org/10.1093/fampra/cm072>.
- Robinson, H. S., and H. Dagfinrud. 2017. "Reliability and Screening Ability of the STarT Back Screening Tool in Patients With Low Back Pain in Physiotherapy Practice, a Cohort Study." *BMC Musculoskeletal Disorders* 18, no. 1: 232. <https://doi.org/10.1186/s12891-017-1553-x>.
- Sharma, S., H. Correia, A. Pathak, et al. 2021. "Translation and Cross-Cultural Adaptation of Nepali Versions of the Patient-Reported Outcomes Measurement Information System (PROMIS®) Pain Intensity, Pain Interference, Pain Behavior, Depression, and Sleep Disturbance Short Forms in Chronic Musculoskeletal Pain." *Quality of Life Research* 30, no. 4: 1215–1224. <https://doi.org/10.1007/s11136-020-02745-6>.
- Sharma, S., M. P. Jensen, G. L. Moseley, and J. H. Abbott. 2018a. "Pain Education for Patients With Non-Specific Low Back Pain in Nepal: Protocol of a Feasibility Randomised Clinical Trial (PEN-LBP Trial)." *BMJ Open* 8, no. 8: e022423. <https://doi.org/10.1136/bmjopen-2018-022423>.
- Sharma, S., M. P. Jensen, G. L. Moseley, and J. H. Abbott. 2019a. "Results of a Feasibility Randomised Clinical Trial on Pain Education for Low Back Pain in Nepal: The Pain Education in Nepal-Low Back Pain (PEN-LBP) Feasibility Trial." *BMJ Open* 9, no. 3: e026874. <https://doi.org/10.1136/bmjopen-2018-026874>.
- Sharma, S., M. P. Jensen, A. Pathak, S. Sharma, M. Pokharel, and J. H. Abbott. 2019b. "State of Clinical Pain Research in Nepal: A Systematic Scoping Review." *Pain Reports* 4, no. 6: e788. <https://doi.org/10.1097/pr9.0000000000000788>.
- Sharma, S., J. Jha, A. Pathak, and R. Neblett. 2020. "Translation, Cross-Cultural Adaptation, and Measurement Properties of the Nepali Version of the Central Sensitization Inventory (CSI)." *BMC Neurology* 20, no. 1: 1–10. <https://doi.org/10.1186/s12883-020-01867-1>.
- Sharma, S., and McAuley, J. H. (2022). "Low Back Pain in Low-And Middle-Income Countries, Part 1: The Problem." *Journal of Orthopaedic and Sports Physical Therapy* 52, no. 5: 233–235. <https://doi.org/10.2519/jospt.2022.11145>.
- Sharma, S., J. Palanchoke, D. Reed, and J. H. Abbott. 2017. "Translation, Cross-Cultural Adaptation and Psychometric Properties of the Nepali Versions of Numerical Pain Rating Scale and Global Rating of Change." *Health and Quality of Life Outcomes* 15, no. 1: 236. <https://doi.org/10.1186/s12955-017-0812-8>.
- Sharma, S., A. Pathak, R. Maharjan, J. H. Abbott, H. Correia, and M. P. Jensen. 2018b. "Psychometric Properties of Nepali Versions of PROMIS Short Form Measures of Pain Intensity, Pain Interference, Pain Behaviour, Depressions, and Sleep Disturbance." *Journal of Pain* 19, no. 3: S59. <https://doi.org/10.1016/j.jpain.2017.12.152>.
- Sharma, S., P. Thibault, J. H. Abbott, and M. P. Jensen. 2018c. "Clinimetric Properties of the Nepali Version of the Pain Catastrophizing Scale in Individuals With Chronic Pain." *Journal of Pain Research* 11: 265–276. <https://doi.org/10.2147/jpr.S153061>.
- Streiner, D. L., G. R. Norman, and J. Cairney. 2024. *Health Measurement Scales: A Practical Guide to Their Development and Use*. Oxford, UK: Oxford University Press.
- Terwee, C. B., S. D. M. Bot, M. R. de Boer, et al. 2007. "Quality Criteria Were Proposed for Measurement Properties of Health Status Questionnaires." *Journal of Clinical Epidemiology* 60, no. 1: 34–42. <https://doi.org/10.1016/j.jclinepi.2006.03.012>.
- Terwee, C. B., C. A. C. Prinsen, A. Chiarotto, et al. 2018. "COSMIN Methodology for Evaluating the Content Validity of Patient-Reported Outcome Measures: A Delphi Study." *Quality of Life Research* 27, no. 5: 1159–1170. <https://doi.org/10.1007/s11136-018-1829-0>.
- Tonosu, J., K. Takeshita, N. Hara, et al. 2012. "The Normative Score and the Cut-Off Value of the Oswestry Disability Index (ODI)." *European Spine Journal* 21, no. 8: 1596–1602. <https://doi.org/10.1007/s00586-012-2173-7>.
- UK, N. G. C. 2016. "National Institute for Health and Care Excellence: Guidelines." In *Low Back Pain and Sciatica in Over 16s: Assessment and Management*. UK: National Institute for Health and Care Excellence. <https://www.ncbi.nlm.nih.gov/books/NBK401577/>.
- Van Wambeke, P., A. Desomer, L. Ailiet, et al. 2017. *Low Back Pain and Radicular Pain: Assessment and Management*. Brussels, Belgium: Belgian Health Care Knowledge Centre (KCE).
- Van Wyngaarden, J. J., B. Noehren, J. S. Pennings, C. Jacobs, P. E. Matuszewski, and K. R. Archer. 2021. "Reliability and Validity Evidence of the STarT-Lower Extremity Screening Tool for Patients With Lower Extremity Fracture: A Prospective Study." *Archives of Physical Medicine and Rehabilitation* 102, no. 2: 261–269. <https://doi.org/10.1016/j.apmr.2020.08.023>.
- Wallwork, S. B., F. A. Braithwaite, M. O'Keeffe, et al. 2024. "The Clinical Course of Acute, Subacute and Persistent Low Back Pain: A Systematic Review and Meta-Analysis." *Canadian Medical Association Journal: Canadian Medical Association Journal* 196, no. 2: E29–E46. <https://doi.org/10.1503/cmaj.230542>.
- Wiangkham, T., N. Phungwattanakul, N. Thongbai, et al. 2021. "Translation, Cross-Cultural Adaptation and Psychometric Validation of the Thai Version of the STarT Back Screening Tool in Patients With Non-Specific Low Back Pain." *BMC Musculoskeletal Disorders* 22, no. 1: 454. <https://doi.org/10.1186/s12891-021-04347-w>.
- Woby, S. R., N. K. Roach, M. Urmston, and P. J. Watson. 2005. "Psychometric Properties of the TSK-11: A Shortened Version of the Tampa Scale for Kinesiophobia." *Pain* 117, no. 1-2: 137–144. <https://doi.org/10.1016/j.pain.2005.05.029>.
- Yılmaz Yelvar, G. D., M. Dalkılıç, Y. Çırak, Y. Parlak Demir, B. N. Karadüz, and M. M. Parlak. 2019. "Validity and Reliability of Turkish Version of STarT Back Screening Tool." *Agri: Agri (Algoloji) Dernegi'nin Yayın organidir = The journal of the Turkish Society of Algology* 31, no. 4: 163–171. <https://doi.org/10.14744/agri.2019.99266>.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.