Validation of the Subgroups for Targeted Treatment for Back (STarT Back) screening tool at a tertiary care centre

Susan Robarts, PT, MSc
Helen Razmjou, PT, PhD
Albert Yee, MD, MSc
Joel Finkelstein, MD, MSc

Correspondence to:
H. Razmjou
Holland Orthopaedic and Arthritic Centre
43 Wellesley St E
Toronto ON M4Y 1H1
helen.razmjou@sunnybrook.ca

Background: The Subgroups for Targeted Treatment for Back (STarT Back) tool is a screening questionnaire developed to identify modifiable risk factors for back pain disability in primary care. Given the ability of this tool to assist with early identification of patients at high risk, we examined its concurrent convergent and known-group construct validity in tertiary care.

Methods: This was a case–control study of adult (age > 18 yr) patients with and without an active work-related compensation claim recruited from an academic health centre between August 2017 and May 2019. Patients in the study group were assessed by a physiotherapist and an orthopedic surgeon in a spine specialty program designed to assess and treat workplace injuries. The control group included patients referred to an orthopedic spine surgeon in a publicly funded specialty clinic where an advanced practice physiotherapist determined the need for surgical consultation. We used the Roland–Morris Disability Questionnaire (RMDQ) and the Hospital Anxiety and Depression Scale (HADS) to determine the convergent and known-group construct validity of the STarT Back tool.

Results: Fifty case and 50 control participants were included. We observed moderate to high association between the STarT Back total score, psychosocial subscore and risk categories and the RMDQ and HADS scores in the expected direction ($p < 0.001$). A significant association was observed between risk group allocation and depression (area under the curve values > 80), having a compensable injury and work status ($p = 0.002$–0.001).

Conclusion: The STarT Back tool was able to differentiate between patients with and without a compensable injury and patients with different levels of work status. The tool has acceptable convergent and known-group construct validity and can assist in clinical decision-making in a tertiary care setting where adjunct psychologic management may be indicated.

Contexte : L'outil STarT Back (pour *Subgroups for Targeted Treatment for Back*) est un questionnaire de triage mis au point pour identifier les facteurs de risque modifiables à l'égard de l'invalidité associée aux dorsalgies en médecine de soins primaires. Étant donné que cet outil peut aider à identifier rapidement les patients à haut risque, nous avons examiné sa validité conceptuelle convergente et discriminante en soins tertiaires.

Méthodes : Il s'agit d'une étude cas–témoins regroupant des patients adultes (âge > 18 ans) ayant ou non en cours une demande d'indemnisation pour accident du travail, recrutés auprès d'un centre universitaire de santé entre août 2017 et mai 2019. Les patients recrutés ont été évalués en physiothérapie et en chirurgie orthopédique dans le cadre d'un programme spécialisé, conçu pour évaluer et traiter les blessures à la colonne vertébrale subies en milieu de travail. Le groupe témoin incluait des patients adressés en chirurgie de la colonne vertébrale dans une clinique publique spécialisée dont le service de physiothérapie en pratique avancée déterminait si une consultation en chirurgie s'imposait. Nous avons utilisé le questionnaire Roland–Morris sur l'invalidité (RMDQ) et l'échelle hospitalière de mesure de l'anxiété et de la dépression (HADS) pour déterminer la validité convergente et discriminante de l'outil STarT Back.

Résultats : Cinquante patients et 50 participants témoins ont été inclus. Nous avons observé un lien de modéré à étroit entre le score STarT Back total, le score à la sous-échelle psychosociale et les catégories de risque selon les scores RMDQ et HADS, selon la tendance prévue ($p < 0,001$). Un lien significatif a été observé entre l'assignation au groupe à risque et la dépression (aires sous la courbe > 80), le fait d'avoir une blessure indemnisable et la situation professionnelle ($p = 0,002$–0,001).

Conclusion : L'outil STarT Back a permis de faire la distinction entre les patients présentant ou non une blessure indemnisable et les patients ayant différentes situations professionnelles. L'outil a une validité conceptuelle convergente et discriminante acceptable et peut aider à la prise de décision clinique dans un contexte de soins tertiaires où une prise en charge psychologique peut être indiquée.

The substantial impact of low back pain on patient function and health care costs has been well established in the literature.[1-4] Although many patient-reported outcome measures have been developed to incorporate patients' perspective and to measure recovery and residual disability, a few have categorized patients into groups according to risk for developing persistent disability.[5,6] The Subgroups for Targeted Treatment for Back (STarT Back) tool is a short screening questionnaire developed in 2008.[6] Its reliability and prognostic utility in primary care patients were investigated in several subsequent studies.[7-10] It has also been validated in multiple languages in primary care and physiotherapy settings.[11-17] The STarT Back tool addresses both physical and psychologic risk factors for persistent disabling back pain and can guide treatment based on risk allocation. For example, it is suggested that patients in the low-risk category receive a 1-time clinic appointment, and those at medium risk be referred to physiotherapy treatment for restoring function and decreasing disabling back or referred leg pain through patient education and reassurance, prescribed physical activity and exercise, manual therapy and acupuncture. The same regimen is recommended for patients at high risk, together with assessment and treatment of biopsychosocial factors. The physical component includes evidence-based exercise protocols and traditional modalities, and psychologically informed techniques are integrated to provide a credible explanation for persistent symptoms, reassurance, education, collaborative goal setting, pacing and graded activity to address patient concerns, maladaptive cognitive behaviours and unhelpful beliefs.

In 2018, the Rapid Access Clinics for Low Back Pain (formerly the Inter-Professional Spine Assessment and Education Clinics), a province-wide program that facilitates management of low back pain within 4 weeks, was established in Ontario, Canada. The Rapid Access Clinics for Low Back Pain adopted the STarT Back tool as part of outcome measurement. In many cases, previous primary care treatment such as medication has not been effective for patients referred to this program, and they require comprehensive assessment, education and treatment provided by community-based advanced practice providers (physiotherapists, chiropractors). When indicated, patients are referred to the hospital-based practice lead (typically an advanced practice physiotherapist), who works in a team approach with the spine surgeon. Validation of the STarT Back tool in these more complex/advanced cases referred to tertiary care is required and will help clinicians use this tool with more confidence.

Given the ability of the STarT Back tool to assist with early identification of patients at high risk, who may benefit from additional structured, psychologically informed interventions, further validation of this tool in tertiary care is warranted. Tertiary care assessment of spine-related problems is often the last resort for patients seeking symptom resolution, and identifying patients who would benefit from or who have not accessed the right treatment package is critical to provide evidence-based care and contain health care costs. The purpose of the present study was to examine the cross-sectional concurrent convergent validity and known-group construct validity of the STarT Back screening tool in patients seen at a tertiary care centre.

METHODS

Design

This study used a case–control design. The study group consisted of people who had experienced a work-related lumbar spine injury and had an active compensation claim. The control group consisted of patients without a work-related spinal injury. We chose the case and control participants based on evidence of a differential level of disability after compensated low back pain.[18-22] We hypothesized that patients with an active compensable injury would show a higher risk of having psychologic concerns, often owing to complex factors such as fear of pain or injury, and more environmental risk factors at the workplace. We recruited both groups from the same academic health centre between August 2017 and May 2019. The study received ethics approval from the Human Ethics Research Board of the Sunnybrook Health Sciences Centre, Toronto (REB# 167-2017). All patients provided informed consent.

Participants

Patients in the study group were assessed by a physiotherapist and an orthopedic surgeon per usual practices in a spine specialty program designed to assess and treat workplace injuries. The program is funded by the Ontario Workplace Safety and Insurance Board, which provides parallel-pay insurance and expedited access to specialist assessment and surgical management.

The control group included patients who were referred to orthopedic spine surgeons in a publicly funded specialty clinic where patients were first seen by an advanced practice physiotherapist to determine the need for surgical consultation.

Inclusion criteria for both groups were age more than 18 years, and ability to write and read English. Exclusion criteria were fracture, infection, chronic pain syndrome, fibromyalgia, diabetic neuropathy or receiving active psychologic intervention. We excluded patients with diabetic neuropathy because its clinical symptomatology overlaps with neurologic symptoms of discogenic origin. We excluded those with fibromyalgia or clinically diagnosed psychologic disorders as these conditions tend to skew the scores of depression, anxiety and pain-related distress, and their management needs a different approach.

Information on participant demographic characteristics, clinical examination and diagnosis was collected by the clinicians.

Screening tool

The STarT Back instrument was conceived primarily as a screening tool for primary care settings, with baseline characteristics as prognostic factors. It consists of 9 items. The first 4 items assess biomedical factors related to referred leg pain, shoulder or neck pain, and inability to walk or dress, and the next 5 items identify modifiable psychosocial risk factors reflecting fear, anxiety, catastrophizing tendency, depression and bothersomeness. A dichotomized response format ("Agree"/"Disagree") is used for the first 8 questions, and a 5-response Likert scale (ranging from "Not at all" to "Extremely") is used for the last question, about bothersomeness. The total score ranges from 0 to 9 and is obtained by summing all positive items. The psychosocial subscale score ranges from 0 to 5 and is obtained by summing the answers to questions 5–9. Patients with a total score of 3 or less are classified as being at low risk for poor prognosis and persistent disability; a total score of 4 points and a score of 3 or less on the psychosocial subscale indicates medium risk; and a score of 4 or more on both the total score and the psychosocial subscale is classified as high risk.

Patient-reported outcome measures used for concurrent validation

We used 2 questionnaires to document disability and anxiety/depression: the Roland–Morris Disability Questionnaire (RMDQ)[23] and the Hospital Anxiety and Depression Scale (HADS).[24] The RMDQ is a self-report disability measure; greater levels of disability are reflected by higher numbers on a 24-point scale. It has established reliability and validity in patients with low back pain.[23,25,26] The HADS measures the extent of mental well-being in relation to anxiety and depression.[24] The total score for both subscales ranges from 0 to 21, with a score of 7 or less for either subscale considered as being in the normal range, 8–10 suggestive of the presence of the mood disorder (borderline), and 11–21 indicating the probable presence of the mood disorder. The HADS has acceptable measurement properties in patients with musculoskeletal conditions[27] and low back pain.[28–30]

Sample size determination

We calculated the sample size for known-group validity, which is more demanding than convergent validity. Assuming that the expected proportion of patients at high risk would be about 30% more in the compensation group than in the noncompensation group, an α level of 0.05 and power of 0.8, we deemed that a minimum of 45 patients per group would be necessary.

Statistical analysis

We examined the convergent validity of the STarT Back tool (total and subscale scores) against the RMDQ and HADS using Spearman rank correlations. We hypothesized a moderate (ρ = 0.5–0.7) association between the continuous data scores of the STarT Back tool and of the RMDQ and HADS. We examined the correlation between the risk categories of the STarT Back tool (low, medium, high) and the HADS categories of normal, borderline and probable presence of the mood disorder using the χ^2 test or Fisher exact test.

We examined the overall ability of the STarT Back risk categories (high v. low/medium) to discriminate between patients with different levels of disability as measured by the RMDQ, and different levels of depression and anxiety as measured by the HADS by plotting receiver operator characteristic (ROC) curves and calculating the areas under curve (AUCs). We used the traditional academic point system as a guide for classifying accuracy (0.90–1.0 = excellent, 0.80–0.89 = good, 0.70–0.79 = fair, and 0.60–0.69 = poor).

We examined the known-group validity by assessing the ability of the STarT Back total score, subscale score and risk categories to differentiate between case and control participants and between different levels of work status (regular/modified full-time, regular/modified part-time, not able to work). Patients who were retired were excluded from this component of the analysis. We hypothesized that patients with a work-related injury would have higher levels of disability and psychosocial concerns and a less desirable work status than patients in the control group.

RESULTS

Fifty patients with a work-related injury and 50 patients without an active compensation claim were included in the study. Patients in the case group were younger than those in the control group and had fewer nontraumatic injuries and less leg-dominant pain, with a lower rate of degenerative disc disease (Table 1). Fewer case than control participants were considered surgical candidates.

Construct convergent validity

As hypothesized, the associations between the total score and subscale score of the STarT Back tool and the scores on the RMDQ and the depression and anxiety components of the HADS were moderate or high, and in the expected direction. The associations between the specific categories of the STarT Back and HADS categories

Table 1. Demographic characteristics of patients with (cases) and without (controls) an active work-related compensation claim

	Group; no. (%) of participants*		
Variable	Case n = 50	Control n = 50	p value†
Age, mean ± SD, yr	40 ± 11	58 ± 16	< 0.001
Sex			0.2
Female	18 (36)	24 (48)	
Male	32 (64)	26 (52)	
Symptom duration, mean ± SD, mo	7 ± 7	31 ± 41	< 0.001
Mechanism of injury			< 0.001
Insidious	5 (10)	45 (90)	
Traumatic	45 (90)	5 (10)	
Type of disorder			
Degenerative disc disease	2 (4)	16 (32)	< 0.001
Disc herniation	12 (24)	16 (32)	0.06
Mechanical low back pain	4 (8)	12 (24)	0.05
Other	32 (64)	6 (12)	—
Pain pattern			
Back dominant	47 (94)	27 (54)	
Leg dominant	3 (6)	23 (46)	< 0.001
Surgical candidate	2 (4)	20 (40)	< 0.001

SD = standard deviation.
*Except where noted otherwise.
†t test, χ² or Fisher exact text, as appropriate.

Fig. 1. Area under the curve values of the receiver operator characteristic (ROC) for disability (Roland–Morris Disability Questionnaire [RMDQ]), depression (Hospital Anxiety and Depression Scale [HADS]) and anxiety (HADS) for the Subgroups for Targeted Treatment for Back screening tool.

were statistically significant for depression ($p < 0.001$) and anxiety ($p < 0.001$), with the majority of patients showing high levels of depression (75 [75%]) and anxiety (62 [62%]) and being classified in the STarT Back high-risk group.

The AUC for the overall ability of the STarT Back risk categories to discriminate between patients with different levels of disability and depression showed that patients in the high-risk group were significantly more disabled than those in the low- or medium-risk group (AUC = 0.89, 95% confidence interval [CI] 0.83–0.96) and had higher levels of depression (AUC = 0.82, 95% CI 0.73–0.91) (Figure 1). The overall ability of the STarT Back risk categories to differentiate between different levels of anxiety was fair (AUC = 0.76, 95% CI 0.66–0.85).

Known-group validity

As hypothesized, the STarT Back total score, subscale score and risk categories were able to differentiate between patients who had an active compensation claim and those who did not (Table 2). The STarT Back total score, subscale score and risk categories were also able to differentiate between different levels of work status (Table 3), with the majority of patients in the nonworking group being classified in the high-risk group (19/28 [68%], v. 11/45 [24%] of full-time workers and 2/9 [22%] of part-time workers).
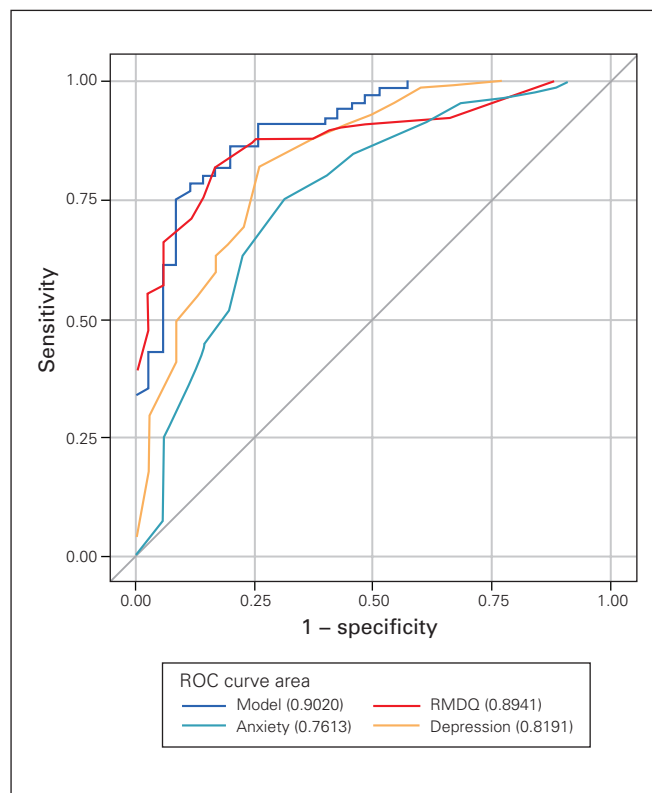
Table 2. STarT Back known-group validity for compensation status

	Group; no. (%) of patients*		
Variable	Case	Control	p value†
Total score, mean ± SD	6.04 ± 2	4.54 ± 2	0.001
Subscale score, mean ± SD	3.32 ± 1	2.26 ± 1	0.001
Risk category			0.002
Low	4 (8)	14 (28)	
Medium	21 (42)	26 (52)	
High	25 (50)	10 (20)	

SD = standard deviation; STarT Back = Subgroups for Targeted Treatment for Back.
*Except where noted otherwise.
†t test or Fisher exact text, as appropriate.

DISCUSSION

In the present study, in a tertiary care setting where patients were referred for surgical consideration for low back pain, the associations between the total score and subscale sore of the STarT Back tool and the scores on the RMDQ and the depression and anxiety components of the HADS were moderate or high, and in the expected direction. The associations between the specific categories of

Table 3. STarT Back known-group validity for work status				
	Work status; no. (%) of patients*†			
Variable	Full-time‡ n = 45	Part-time‡ n = 9	Not working n = 28	p value§
Total score, mean ± SD	4.8 ± 2	5.22 ± 2	6.79 ± 1	0.001
Subscale score, mean ± SD	2.47 ± 2	2.56 ± 1	3.75 ± 1	0.001
Risk category				0.002
Low	11 (24)	1 (11)	1 (4)	
Medium	23 (51)	6 (67)	8 (29)	
High	11 (24)	2 (22)	19 (68)	

SD = standard deviation; STarT Back = Subgroups for Targeted Treatment for Back.
*Except where noted otherwise.
†Eighteen patients were on training or retired.
‡Regular/modified.
§F statistic or Fisher exact test, as appropriate.

the STarT Back and HADS categories were statistically significant for depression and anxiety. The STarT Back tool was able to discriminate between groups with and without a compensation claim, a factor that is known to affect disability and recovery.[18–22] In a study conducted in the United Kingdom, the investigators compared the STarT Back tool and the Örebro Musculoskeletal Pain Screening Questionnaire[10] and reported similarities between the 2 measures, with moderate agreement between the 2 (weighted κ = 0.57). The Spearman rank correlation coefficients for the total score and psychosocial subscale score were 0.80 and 0.77 with the Örebro Musculoskeletal Pain Screening Questionnaire scores, respectively. The correlation coefficient between the psychosocial subscale score and the RMDQ score was 0.81.

In a randomized controlled trial conducted in the UK, Hill and colleagues[31] randomly assigned 1573 participants to the STarT Back guided intervention (study group) or the control group, with the RMDQ being used as the primary outcome score at 12 months. They found that stratified care was associated with a mean increase in generic health benefit and cost savings (£240.01 [about Can$388 in 2022 dollars] v. £274.40 [about Can$443]) compared to the control group. Patients in the low-risk group received a 1-time assessment, with education and reassurance that further treatment was unlikely to be necessary. Patients in the medium- and high-risk groups received 6 sessions over 3 months focusing on education and evidence-based treatment, together with psychologically informed management for the high-risk group after referred leg pain/radiculopathy was ruled out. The role of specific prognostic psychologic indicators identified by the STarT Back tool (i.e., low mood, anxiety, pain-related fear and catastrophizing) was addressed in the high-risk group. The authors concluded that systematic screening with the STarT Back tool could assist decision-making and treatment referrals.

In a study conducted in primary care in the United States, Suri and colleagues[7] reported that the STarT Back risk groups successfully classified patients with back pain into distinct categories of risk for persistent disabling back pain at 6 months. Beneciuk and colleagues[9] investigated the predictive validity of the STarT Back tool in predicting 6-month clinical outcomes in an outpatient physiotherapy setting and found that both baseline and 4-week scores were valuable in predicting 6-month Oswestry scores. In a subsequent article, they suggested that a 2-group risk category might provide a clearer representation of the level of pain-associated psychologic distress, maladaptive coping and disability in the outpatient physiotherapy setting.[8] We found that both the 2-group (high v. low/medium) and the 3-group (high, medium, low) stratified classification categories were valid in differentiating between levels of disability, as measured by the RMDQ, and of mental well-being, as measured by the HADS. In a study conducted in Australia, the STarT Back tool provided an acceptable indication of 1-year disability but had poor predictive and discriminative ability for future pain in a population with chronic low back pain.[32]

Studies that have used a translated version of the STarT Back tool in primary care or rehabilitation settings have shown promising results. Robinson and Dagfinrud[11] explored the reliability and screening ability of the STarT Back tool in a physiotherapy clinic in Norway and found that it was reliable and able to stratify patients into risk groups. French investigators reported a high Spearman correlation coefficient (0.74) between the STarT Back tool and the RMDQ[12] in primary care. In a study similar to ours, Abedi and colleagues[17] reported a high correlation (> 0.70) between the Persian version of the STarT Back tool and the RMDQ (0.081) and the 2 subscales of the HADS. In the present study, the correlation coefficients were 0.75 with the RMDQ, 0.67 with the depression component of the HADS and 0.50 with the anxiety component of the HADS. The fact that our sample was more diverse may explain the slightly lower values. The total and psychosocial subscale scores of the Brazilian version of the STarT Back tool were reported to have good correlation with the RMDQ (r = 0.70 and r = 0.64, respectively).[13] The AUC for the discriminant validity of the total and psychosocial subscale scores against the reference standard was 0.88 for disability,[13] similar to the value in our study (0.89). The German version of the STarT Back tool showed an AUC of 0.79 for discriminating chronic pain status at 12 months.[16] Finally, Forsbrand and colleagues[15] examined the predictive ability of the STarT Back risk groups in relation to health-related quality of life and work ability at follow-up in southern Sweden. Patients in the high-risk group had a significantly increased risk of having poor health-related quality of life and poor work

ability at a median of 13 months. The AUC was 0.73 for health-related quality of life and 0.68 for work ability.

Limitations

This study has limitations owing to its cross-sectional nature, which did not allow assessment of longitudinal predictive validity. Generalizability of validity studies is limited to patients with similar characteristics, and our results are applicable to injured workers with acute injury and noninjured workers referred to a specialty spine clinic in an academic health centre. Providing evidence for cross-sectional convergent and known-group construct validity of the STarT Back tool is the first step in validating this instrument for use in specialty spine clinics, but further validation of the tool is recommended in longitudinal studies and in patients with different levels of spinal disorders.

CONCLUSION

The STarT Back screening tool has acceptable convergent and known-group validity and has the potential to identify patients who may benefit from adjunct psychologic management. This is important to clinicians in tertiary care, where patients and their primary care providers are seeking definitive solutions to spine-related complaints. Systematic screening for maladaptive cognitive, psychologic or social factors will assist in referring patients to the right treatments and help contain costs by diverting those at high risk for persistent disability from care and additional diagnostic testing that is not likely to be of value.

Affiliations: From the Department of Rehabilitation, Holland Orthopedic & Arthritic Centre, Sunnybrook Health Sciences Centre, Toronto, Ont. (Robarts, Razmjou); the Department of Physical Therapy, Faculty of Medicine, University of Toronto, Toronto, Ont. (Robarts, Razmjou); Sunnybrook Research Institute, Sunnybrook Health Sciences Centre, Toronto, Ont. (Robarts, Razmjou, Yee, Finkelstein); the Division of Orthopedic Surgery, Department of Surgery, Sunnybrook Health Sciences Centre, Toronto, Ont. (Yee, Finkelstein); and the Division of Orthopaedic Surgery, Department of Surgery, Faculty of Medicine, University of Toronto, Toronto, Ont. (Yee, Finkelstein).

References

1. Global Burden of Disease Study 2013 Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* 2015;386:743-800.
2. Lim KL, Jacobs P, Klarenbach S. A population-based analysis of healthcare utilization of persons with back disorders: results from the Canadian Community Health Survey 2000–2001. *Spine (Phila Pa 1976)* 2006;31:212-8.
3. Katz JN. Lumbar disc disorders and low-back pain: socioeconomic factors and consequences. *J Bone Joint Surg Am* 2006;88(Suppl 2):21-4.
4. Gandjour A, Telzerow A, Lauterbach KW; INTERCARE International Investigators. European comparison of costs and quality in the treatment of acute back pain. *Spine (Phila Pa 1976)* 2005;30:969-75.
5. Gabel CP, Burkett B, Melloh M. The shortened Örebro Musculoskeletal Screening Questionnaire: evaluation in a work-injured population. *Man Ther* 2013;18:378-85.
6. Hill JC, Dunn KM, Lewis M, et al. A primary care back pain screening tool: identifying patient subgroups for initial treatment. *Arthritis Rheum* 2008;59:632-41.
7. Suri P, Delaney K, Rundell SD, et al. Predictive validity of the STarT Back tool for risk of persistent disabling back pain in a U.S. primary care setting. *Arch Phys Med Rehabil* 2018;99: 1533-9.e2.
8. Beneciuk JM, Robinson ME, George SZ. Subgrouping for patients with low back pain: a multidimensional approach incorporating cluster analysis and the STarT Back Screening Tool. *J Pain* 2015;16:19-30.
9. Beneciuk JM, Bishop MD, Fritz JM, et al. The STarT Back Screening Tool and individual psychological measures: evaluation of prognostic capabilities for low-back pain clinical outcomes in outpatient physical therapy settings. *Phys Ther* 2013;93:321-33.
10. Hill JC, Dunn KM, Main CJ, et al. Subgrouping low back pain: a comparison of the STarT Back Tool with the Orebro Musculoskeletal Pain Screening Questionnaire. *Eur J Pain* 2010;14:83-9.
11. Robinson HS, Dagfinrud H. Reliability and screening ability of the STarT Back screening tool in patients with low back pain in physiotherapy practice, a cohort study. *BMC Musculoskelet Disord* 2017;18:232.
12. Bruyère O, Demoulin M, Beaudart C, et al. Validity and reliability of the French version of the STarT Back Screening Tool for patients with low back pain. *Spine (Phila Pa 1976)* 2014;39:E123-8.
13. Pilz B, Vasconcelos RA, Teixeira PP, et al. Construct and discriminant validity of STarT Back Screening Tool — Brazilian version. *Braz J Phys Ther* 2017;21:69-73.
14. Forsbrand M, Grahn B, Hill JC, et al. Comparison of the Swedish STarT Back Screening Tool and the short form of the Örebro Musculoskeletal Pain Screening Questionnaire in patients with acute or subacute back and neck pain. *BMC Musculoskelet Disord* 2017;18:89.
15. Forsbrand MH, Grahn B, Hill JC, et al. Can the STarT Back Tool predict health-related quality of life and work ability after an acute/subacute episode with back or neck pain? A psychometric validation study in primary care. *BMJ Open* 2018;8:e021748.
16. Karstens S, Krug K, Raspe H, et al. Prognostic ability of the German version of the STarT Back Tool: analysis of 12-month follow-up data from a randomized controlled trial. *BMC Musculoskelet Disord* 2019;20:94.
17. 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;20:850-4.
18. Carnide N, Hogg-Johnson S, Koehoorn M, et al. Relationship between early prescription dispensing patterns and work disability in a cohort of low back pain workers' compensation claimants: a historical cohort study. *Occup Environ Med* 2019;76:573-81.

19. Fernandez M, Moore C, Peng W, et al. The profile of chiropractors managing patients with low back-related leg pain: analyses of 1907 chiropractors from the ACORN practice-based research network. *Chiropr Man Therap* 2019;27:19.

20. Hanvold TN, Kines P, Nykanen M, et al. Occupational safety and health among young workers in the Nordic countries: a systematic literature review. *Saf Health Work* 2019;10:3-20.

21. Lee YP, Farhan SD, Kiester D, et al. Variables affecting return to work after spinal surgery in a non-workers' compensation population: a retrospective cohort study. *J Am Acad Orthop Surg* 2017;25:e282-8.

22. Tye EY, Anderson JT, Haas AR, et al. The timing of surgery affects return to work rates in patients with degenerative lumbar stenosis in a workers' compensation setting. *Clin Spine Surg* 2017;30:E1444-9.

23. 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;8:145-50.

24. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand* 1983;67:361-70.

25. Chiarotto A, Ostelo RW, Boers M, et al. A systematic review highlights the need to investigate the content validity of patient-reported outcome measures for physical functioning in patients with low back pain. *J Clin Epidemiol* 2018;95:73-93.

26. Yao M, Zhu S, Tian ZR, et al. Cross-cultural adaptation of Roland–Morris Disability Questionnaire needs to assess the measurement properties: a systematic review. *J Clin Epidemiol* 2018;99:113-22.

27. Pallant JF, Tennant A. An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS). *Br J Clin Psychol* 2007;46(pt 1):1-18.

28. Robarts S, Razmjou H, Yee A, et al. Psychometric properties of the Optimal Screening for Prediction of Referral and Outcome Yellow Flags (OSPRO-YF) in patients with lumbar spine pain. *Eur J Physiother* 2021;23:234-40.

29. Razmjou H. Psychosocial flag signs in patients with compensable occupational lumbar spine injuries. *J Orthopaed Physiother* 2018;1:1-7.

30. Carreon LY, Jespersen AB, Stottrup CC, et al. Is the Hospital Anxiety and Depression Scale associated with outcomes after lumbar spine surgery? *Global Spine J* 2020;10:266-71.

31. Hill JC, Whitehurst DGT, Lewis M, et al. Comparison of stratified primary care management for low back pain with current best practice (STarT Back): a randomised controlled trial. *Lancet* 2011;378:1560-71.

32. Kendell M, Beales D, O'Sullivan P, et al. The predictive ability of the STarT Back Tool was limited in people with chronic low back pain: a prospective cohort study. *J Physiother* 2018;64:107-13.