Comparing the STarT Back Screening Tool's Subgroup Allocation of Individual Patients With That of Independent Clinical Experts

Jonathan C. Hill, PhD, Kanchan Vohora, BSc, Kate M. Dunn, PhD, Chris J. Main, PhD, and Elaine M. Hay, MD

Objectives: The STarT Back Screening Tool (SBST) is validated to subgroup primary care patients with back pain into risk groups relevant to initial decision-making. However, it remains unclear how the tool's allocation of individuals compares with subjective clinical decision-making. We evaluated agreement between clinicians and the SBST's allocation to risk subgroups, and explored reasons for differences observed.

Methods: Twelve primary care back pain patients underwent a video-recorded clinical assessment. The SBST was completed on the same day. Clinical experts (3 general practitioners, 3 physiotherapists, and 3 pain management specialists) individually reviewed the patient videos (4 each), blind to SBST allocation. Their task was to subgroup patients into low, medium, or high-risk groups.

Results: Interrater agreement between clinicians was "fair" ($\kappa = 0.28$), with consistent allocation between experts in 4 of 12 patients. There was observed agreement with the SBST in 17 of 36 cases (47%) and Cohen's weighted κ was 0.22, indicating fair agreement. Two reasons for differences emerged. Clinicians tailor their decisions according to patient expectations and demands for treatment and clinicians use knowledge of difficult life circumstances that may be unrelated back pain.

Discussion: Clinicians make inconsistent risk estimations for primary care patients with back pain when using intuition alone, with little agreement with a formal subgrouping tool. Unlike clinicians, the SBST could not make a sophisticated synthesis of patient preferences, expectations, and previous treatment history. Although acknowledging the limitations of back pain subgrouping tools, more research is needed to test whether their use improves consistency in primary care decision-making.

Key Words: low back pain, subgroups, primary care, clinical decision making

(*Clin J Pain* 2010;26:783–787)

Primary care practitioners frequently make difficult clinical decisions about the management of patients with nonspecific low back pain (LBP).[1] With the lack of a reliable biomedical diagnosis to guide treatment options, evidence-based guidelines recommend clinicians consider the presence of psychosocial prognostic indicators as part of their decision-making process.[2,3] To help practitioners identify important prognostic indicators, formal prognostic tools have been developed that provide recommended cut-off scores to facilitate decision-making about initial treatment options based on likely prognosis.[4–6] However, it remains unclear how such tools compare with clinical decision-making performed using clinicians' subjective intuition alone.

The STarT Back Screening Tool (SBST) is a brief, newly validated tool[7] designed to screen primary care patients with LBP for prognostic indicators that are relevant to initial decision-making. Previous research has examined the SBST in a head-to-head comparison with the Örebro Musculoskeletal Pain Screening Questionnaire[4] and tested its predictive validity within an external sample.[7] The instrument is designed for use in clinical practice to help clinicians systematically identify patients "at risk" for persistent LBP symptoms. Validated cut-off scores are available to help clinicians allocate individual patients into 1 of 3 initial treatment options (low, medium, and high-risk subgroups). A randomized clinical trial is currently in progress to provide answers about whether a subgrouping and targeted treatment approach to the management of primary care LBP produces improved patient outcomes compared with clinical decision-making without using the tool and linked treatment pathways.[8] However, this clinical trial is not designed to compare the SBST's subgroup allocation of individual patients with that of independent clinical experts.

Previous studies have compared clinician identification of patients with particular characteristics, and the results have been inconsistent. Two studies of the ability of clinicians to identify specific psychological constructs such as depression during their subjective assessment concluded that clinician identification is poor in comparison with formal depression screening tools.[9,10] In contrast, research comparing the intuitive abilities of general practitioners (GPs) to predict patient outcomes against the predictive abilities of statistically derived clinical prediction rules within primary care LBP patients[11] reported that risk estimation by GPs was only marginally improved by formal prediction rules. Without clinical trial evidence, the question, therefore, remains about the usefulness of adopting formal prediction rules into clinical practice.

This study was designed to (1) enable a comparative evaluation of interrater agreement between clinical experts' intuitive allocation of individual patients to low, medium, and high-risk estimation subgroups, and agreement between

Received for publication April 28, 2010; revised May 25, 2010; accepted May 29, 2010.
From the Arthritis Research UK Primary Care Centre, Keele University, Keele, Staffordshire, UK.
Supported by a program grant from Arthritis Research UK Primary Care Centre, Keele University, Keele, Staffordshire, UK and by the North Staffordshire Primary Care Research Consortium, Keele University, Keele, Staffordshire, UK. Dr Hill holds an Arthritis Research UK Lectureship in Physiotherapy (5 y award).
Reprints: Jonathan C. Hill, PhD, Lecturer in Physiotherapy, Arthritis Research UK Primary Care Centre, Keele University, Keele, Staffordshire, ST5 5BG, UK (e-mail: j.hill@cphc.keele.ac.uk).
Copyright © 2010 by Lippincott Williams & Wilkins

experts and the SBST's subgroup allocation and (2) to explore the reasons for any differences observed.

MATERIALS AND METHODS

A patient clinical assessment study was designed to enable the agreement between the screening tool and experts' subjective risk allocation. Ethical approval to perform the study was obtained from the North Staffordshire Local Research Ethics Committee (Project Number 04/Q2604/116).

Participants and Setting

Primary care LBP consecutive consulters participating in a cross-sectional study (n = 244) from March to April 2005[7] were invited to take part in this study. Details of that study are presented elsewhere,[12] but in brief patients presenting to their GP with "nonspecific" LBP from 8 UK GP practices aged 18 to 59 years were invited to complete a questionnaire containing the SBST. Responders to the survey, who consented to further follow up, were invited by telephone to participate in a clinical assessment study, providing they continued to have significant LBP symptoms and were happy to undergo a video-recorded clinical assessment. This was carried out by the research nurse receiving the questionnaires, using the SBST scores to purposively sample the first 4 consecutive consenting patients from each of the SBST's 3 prognostic risk subgroups. Details of the SBST, the multifactorial items it contains, its methods of scoring, and a number of different translations are freely available at www.keele.ac.uk/startback.

Clinical Assessment

Clinical assessments were conducted over a 3-weeks period by an experienced clinician with expertise in back pain. Their purpose was to elicit clinical information to enable clinical experts reviewing a video of the consultation to allocate patients into a prognostic subgroup (low, medium, or high risk). Immediately before the assessment patients completed the SBST, and were allocated to 1 of its 3 prognostic subgroups for comparison with clinical experts. The assessments were conducted within a local medical research unit, taking an average of 30 minutes. The content of the assessment was based on a published format[13] which encourages patients to disclose pertinent biomedical details about their back problem as well as their beliefs, concerns, and worries (Fig. 1). The video recording equipment was set up beforehand so that no operator was present during the clinical assessment.

Nine clinical experts from 3 professional groups who manage LBP in primary care were invited to participate in the study, including 3 GPs, 3 physiotherapists, and 3 pain management specialists. These experts were independent of the research team and unaware of the questions included in the screening tool. Their task was to each individually review 4 of the patient assessment videos, while blinded to the results of the SBST, and to allocate each patient to 1 of the 3 risk subgroups and state their reasons for allocation. Therefore each of the 12 patient videos was reviewed by 1 person from each of the 3 professional groups represented. The 3 subgroups were described to the clinicians using a standardized written explanation as follows:

1. Allocate the patient to the low-risk subgroup if they appear from the assessment to be at a low risk of persistent, disabling pain, and could be appropriately managed with reassurance, advice, and analgesia by their GP alone.
2. Allocate the patient to the medium-risk subgroup if they appear from the assessment to display significant physical risk factors, but are not displaying severe psychosocial risk factors, and are in their opinion likely to be successfully managed by a physical therapist.
3. Allocate the patient to the high-risk subgroup if they appear from the assessment to display significant psychosocial factors (with or without physical risk factors) that are likely to be most appropriately managed by a clinician with competency to treat psychosocial obstacles to recovery.

The experts then rated their confidence in their decision-making (using a Likert scale from "very confident" to "not at all confident") for each patient and provided additional comments to justify their subgroup allocation.

Analysis of Data

The demographic and clinical characteristics of the patient sample were captured from the results of the survey questionnaire completed 3 to 6 weeks before the assessment study. Differences between the patients selected for assessment and the whole survey sample were described to determine whether they were representative of the whole sample by calculating median scores for relevant clinical characteristics (mean values were used for age).

Observed agreement was examined between the following:

(1) the allocation from the screening tool and the allocation of the experts' to the 3 subgroups (eg, 12 patients each with 3 clinical expert reviews, n = 36 decisions)
(2) consensus expert opinion (consensus was defined where 2 or 3 experts allocated the patient to the same subgroup) and the screening tool's subgroup allocation
(3) interrater agreement between the clinical experts.

Agreement beyond chance for each of these comparisons was statistically evaluated using a weighted Cohen's κ test. The κ values were classified for reference as follows: < 0.00 showed poor agreement; 0.00 to 0.20, slight agreement; 0.21 to 0.40, fair agreement; 0.41 to 0.60, moderate agreement; 0.61 to 0.80, substantial agreement; and greater than 0.80, near perfect agreement.[14] The confidence of each clinician's decision was compared against the tool's subgroup allocation to determine if clinicians were more confident in allocating patients to low, medium, or high-risk groups.

The experts' statements regarding their reasons for subgroup allocation were then descriptively analyzed to explore possible reasons for differences between experts and the screening tool. To facilitate this analysis, clinical characteristics and clinicians statements were examined side-by-side to elicit reasons for disagreements.

RESULTS

Patient Sample Characteristics

The mean age of the 12 patients assessed was 45.2 years [compared with (cf) 44.1 y in the survey sample] ranging from 29 to 58 years, 83% (cf 58%) were female. One of the patients was currently off-work owing to his back pain, and the median number of days taken off-work

© 2010 Lippincott Williams & Wilkins

Interview Schedule

Key themes to be discussed with patients.

A: Presenting characteristics

1 Presenting pain problem: location, severity, duration, type of onset
2 Clinical history: pain, general health, psychological problems
3 Treatment history: diagnostic formulations and emotional reactions
4 Current occupational and social circumstances

B: Specific Stem Dimensions

5 Attitudes and beliefs about pain: cause, control, hurt & harm, outcome
6 Behaviours: description of current coping strategies
7 Financial impact: Work loss and job threat, benefits and litigation
8 Diagnosis and treatment issues: iatrogenic confusion & distress
9 Emotional impact: Anxiety and fears about diagnosis, pain associated limitations, pain persistence, mood (irritability, anger & depression), sleep and fatigue
10 Family: practical and emotional impact on family, reactions of family
11 Work: current and anticipated impact on work, job description (demand, stress, satisfaction), perceived obstacles to work (physical, logistical, social climate)

(Source Reference - Main and Watson (2001) pp 175-200.)

Additional biomedical questions to ask at interview

1. Do you have back pain at present? Are symptoms improving, unchanging or worsening?
2. When did your back problems start this time around and what was the cause?
3. What appears to make the symptoms worse and what makes symptoms better?
4. Do your symptoms disturb your sleep?
5. Can you stand up and show me where you feel your pain?
6. Does your pain spread at all down your legs?
7. Is the pain there all the time?
8. Could you describe the type of pain you feel?
9. Do you have difficulty moving after staying in one position too long because of your back?
10. Have you ever had other joint problems apart from your back?
11. Since your back pain started: have you had any difficulty passing urine? Have you had any problems with your walking or steadiness on your feet? Have you had any unexplained weight loss?
12. Have you had previous episodes of similar symptoms?
13. Have you ever been diagnosed with any medical conditions?
14. Have you ever had any operations to your back?
15. Have any of the treatments you have tried worked well?
16. Have you seen a doctor at the hospital about your back problem?
17. Have you seen any other health care professional then your GP about your back problem?
18. Are you waiting for any appointments/treatments for your back pain? If so who?
19. Has you doctor prescribed any medications or do you take any over-the-counter medication's?
20. What do you believe is the matter with your back?
21. What would you consider to be your two most important health problems at the moment?
22. What treatments are you trying at the moment?
23. Does your pain limit your activities?

FIGURE 1. Assessment schedule.

owing to back pain in the last 6-months was "1 to 7 days." Four patients reported having had their back pain for between 1 month and 3 years; 4 patients had durations of <1 month and 4 had durations of more than 3 years. The clinical characteristics of the 12 assessment sample patients were similar to the cross-sectional survey sample (n = 131) (Hill et al, 2010) with median scores for the screening tool, pain intensity, fear of movement (Tampa Scale for Kinesiophobia), and function (Roland Morris Disability Questionnaire), of 5.0 (cf 5.0), 3.8 (cf 3.7), 38.0 (cf 41.0), and 10.0 (cf 7.0), respectively.

Observed Agreement Between the Screening Tool and Experts

On the day of clinical assessment, patients' subgroup allocations by the SBST were: 8 low risk, 3 medium risk, and 1 high risk. This differed from their original subgroup characteristics when identified by their postal questionnaire, probably owing to the time lag of between 3 to 6 weeks. The 9 clinical experts allocated the patients 36 times as each patient was allocated independently 3 times; 15 patients as low risk, 12 as medium risk, and 9 as high risk. The interrater agreement between clinicians' subgroup allocation was poor, with consistent allocation by all 3 experts in only 4 of 12 patients. The observed agreement between the screening tool and clinical experts in subgroup allocation was 17 of 36 decisions (47%). Table 1 provides a cross-tabulation of screening tool allocation against the expert's allocation for each of the 36 decisions.

The Cohen's weighted κ for the agreement between the tool and experts in subgroup allocation was 0.22 [95% confidence interval (CI) 0.02, 0.42], indicating fair agreement. The κ for the interrater agreement was 0.28 ($P = 0.010$, SE = 0.119), indicating fair agreement. Different

© 2010 Lippincott Williams & Wilkins

TABLE 1. Cross-tabulation of Tool Allocation Against Expert Allocation

	Screening Tool			
	Low Risk	Medium Risk	High Risk	Total
Expert				
Low risk	13	2	0	15
Medium risk	7	3	2	12
High risk	4	4	1	9
Total	24	9	3	36

professional groups demonstrated different levels of agreement with the screening tool with a weighted κ among GPs, pain team, and physiotherapists of 0.00 (95% CI −0.25, 0.25), 0.33 (95% CI 0.02, 0.65), and 0.37 (−0.06, 0.80), respectively. Agreement (weighted κ) between expert clinicians' consensus opinion and the screening tool was 6 of 12 (0.25, 95% CI −0.17-0.67) indicating fair agreement. Among patients allocated by the tool as low risk, clinician consensus agreed in 5 of 8 patients, whereas among the tool's medium-risk group agreement with clinicians was in 1 of 3 patients. It was noted that overall the expert thresholds for subgroup allocation were not consistently lower, nor higher than the subgroup allocation made by the SBST, and discrepancy between the SBST and experts was greatest among patients whose SBST scores were close to cut-off thresholds. Clinical experts' confidence in allocation to subgroup was highest for allocation to the low-risk group and decreased with allocation to medium and high-risk groups (as presented in Table 2).

Reasons for Lack of Agreement

Clinical experts' statements of their reasons for risk subgroup allocation for specific patients were explored together with the patient characteristics to try to explain the poor agreement observed between the clinical experts and the screening tool. Two specific conclusions were drawn from this process. First, that the experts tailored their subgroup recommendations according to their understanding of the patient's expectations and demands for treatment. For example, 2 patients allocated as low-risk by the SBST were allocated to the high-risk subgroup by experts because they had a strong desire for treatment which was communicated during the clinical assessment. Second, the experts identified generally difficult life circumstances, whereas the SBST only identified back pain-related distress. For example, a GP reviewer allocated a patient to the high-risk subgroup because: "this patient does not feel as if he has any 'real friends' and his home situation maybe stressful living with a sick mother-in-law."

TABLE 2. Cross-tabulation of Expert Allocation Against Confidence Ratings

	Confidence Rating			
	Very Confident	Quite Confident	Neither	Total
Expert				
Low risk	6	9	0	15
Medium risk	2	10	0	12
High risk	3	4	2	9
Total	11	23	2	36

DISCUSSION

This study evaluated the comparative agreement between clinical experts' intuitive allocation and a formal screening tool's allocation of individual back pain patients to low, medium, and high-risk estimation subgroups, and explored the reasons for any differences observed. The results revealed that agreement between clinical experts and a formal screening tool occurred in only half of the study patients, which was only a little better than chance (a third). The findings also suggest that clinicians make very inconsistent treatment decisions for patients with back pain presenting to primary care when using their clinical acumen alone. The low interrater agreement between the clinicians together with the fact that clinicians' confidence in allocation to subgroup decreased among patients with the worst prognosis, suggest that some formal assistance to standardize clinical decision-making, particularly for more complex patient problems may prove beneficial.

However, such formal clinical subgrouping tools are not themselves without their own faults and weaknesses. For example, the SBST failed in a number of cases to identify nonpain-related psychosocial distress and difficult life circumstances that the clinicians were able to identify from the assessment and used to allocate patients to the high-risk subgroup because of their potential influence on the persistence of symptoms. The SBST was also not able to make a sophisticated synthesis of patient preferences, expectations, and previous treatment history, despite the obvious importance of these factors to clinicians in their decision-making.

The strengths and weaknesses of the SBST's subgrouping abilities have important implications for practice. Clinicians using the SBST to help identify back pain subgroups need to retain clinical caution and not unquestioningly use cut-off recommendations. The tool uses validated cut-off thresholds based on average group effects, which do not always ensure appropriate subgroup allocation for individual patients. Clinicians' are skilled in operating across the belief systems of patients, including their experiences of pain, ideas about causality, previous healthcare encounters, comorbidity, and life-world experiences.[15] In practice clinicians are, therefore, advised to use the instrument as an adjunct to their own decision-making rather than a replacement to their considered clinical acumen. The strengths of the SBST are likely to be its systematic and consistent allocation of patients to subgroups, which contrasts with the experts' overwhelming inconsistencies in decision-making, and their lack of confidence in decision-making among more complex back pain cases. It seems reasonable to hypothesize that for the more complex patients where the clinicians' confidence in decision-making is less, the tool may act to serve as a useful method of directing a clinician's assessment about a patient's concerns, expectations, and mood, particularly during brief triage focused primary care consultations. An ongoing clinical trial comparing the clinical outcomes of using the SBST's subgrouping and targeted treatment approach is currently in progress and shortly owing to report its findings.[8]

There were a number of methodological limitations to the study design. The selection process did not succeed in identifying an equal spread of patients across the 3

© 2010 Lippincott Williams & Wilkins

screening tool subgroups, and only one patient was identified as high-risk on the day of the video assessment. This limited the understanding of decision-making for patients with high pain-related psychosocial distress. The difficulties in recruiting high-risk patients reflects the natural history of the early stages of acute LBP, with the majority of patients experiencing progress over the first few weeks after consultation.[16] In future similar studies, it is recommended that assessments are conducted closer to the initial data collection time point. A further limitation was that the clinicians were not asked to rereview patient video tapes a second time, with screening tool subgroup allocation available to them. This would have directly enabled an investigation of whether knowledge of the screening tool's allocation changed their subgroup allocation. In addition, the number of patients and clinicians involved in the study was limited, which may influence the generalizability of the findings and it is, therefore, recommended that future studies include more patients and practitioners.

In summary, this study demonstrated poor agreement between clinical experts and the SBST's allocation to LBP subgroups, owing to a combination of factors related to the strengths and weaknesses of brief formal clinical tools. The findings also highlighted the difficulties clinicians have in making consistent and confident clinical decisions about the management of complex individuals with LBP presenting to primary care. Implications for practice include the need for caution in using the SBST to determine subgroup allocation without being interpreted within the context of a clinical assessment, particularly where considerations regarding previous treatment experiences, patient preferences, and difficult life circumstances are important to the decision-making process. Further research is in progress to evaluate whether the SBST provides better clinical outcomes when it is integrated into primary care clinical practice.

ACKNOWLEDGMENTS

The authors thank the administrative and health informatics staff at Keele University's Primary Care Musculoskeletal Research Centre and the Keele General Practice Partnership. They also thank the wider members of the research team who were involved with the study and the 9 expert clinicians who took part: George Peat, PhD (Arthritis Research UK Primary Care Centre, Keele University, Keele, Staffordshire, UK), Sarah Ryan, PhD (Haywood Hospital, High Lane, Burslem, Stoke on Trent, UK), Kay Stevenson, MPhil (University Hospital of North Staffordshire, Stoke on Trent, Staffordshire, UK), Paul Watson, PhD (Department of Health Sciences, New Academic Unit, Leicester General Hospital, Leicester, UK), Mark Porcheret, MD (Arthritis Research UK Primary Care Centre, Keele University, Keele, Staffordshire, UK), Hannah Morphy, MBChB (Arthritis Research UK Primary Care Centre, Keele University, Keele, Staffordshire, UK), Andrew Holdom, BSc (Focus on Backs Pty Ltd, Crows Nest, New South Wales, Australia), Laurence Wood, PhD (Arthritis Research UK Primary Care Centre, Keele University, Keele, Staffordshire, UK), and Christian Mallen, PhD (Arthritis Research UK Primary Care Centre, Keele University, Keele, Staffordshire, UK).

REFERENCES

1. Corbett M, Foster N, Ong BN. GP attitudes and self-reported behaviour in primary care consultations for low back pain. *Fam Pract*. 2009;26:359–364.
2. Chou R, Qaseem A, Snow V, et al. Diagnosis and treatment of low back pain: a joint clinical practice guideline from the American College of Physicians and the American Pain Society. *Ann Intern Med*. 2007;147:478–491.
3. Van Tulder M, Becker A, Bekkering T, et al. Chapter 3. European guidelines for the management of acute nonspecific low back pain in primary care. *Eur Spine J*. 2006;15(suppl 2):S169–S191.
4. Linton SJ, Hallden K. Can we screen for problematic back pain? A screening questionnaire for predicting outcome in acute and subacute back pain. *Clin J Pain*. 1998;14:209–215.
5. Linton SJ, Boersma K. Early identification of patients at risk of developing a persistent back problem: the predictive validity of the Örebro Musculoskeletal Pain Questionnaire. *Clin J Pain*. 2003;19:80–86.
6. Denison E, Asenlöf P, Sandborgh M, et al. Musculoskeletal pain in primary health care: subgroups based on pain intensity, disability, self-efficacy, and fear-avoidance variables. *J Pain*. 2007;8:67–74.
7. Hill JC, Dunn KM, Lewis M, et al. A primary care back pain screening tool: identifying patient subgroups for initial treatment. *Arthritis Rheum*. 2008;59:632–641.
8. Hay EM, Dunn KM, Hill JC, et al. A randomised clinical trial of subgrouping and targeted treatment for low back pain compared with best current care. The STarT Back Trial Study Protocol. *BMC Musculoskelet Disord*. 2008;9:58.
9. Haggman S, Maher CG, Refshauge KM. Screening for symptoms of depression by physical therapists managing low back pain. *Phys Ther*. 2004;84:1157–1166.
10. Spitzer RL, Williams JB, Kroenke K, et al. Utility of a new procedure for diagnosing mental disorders in primary care. The PRIME-MD 1000 study. *JAMA*. 1994;272:1749–1756.
11. Jellema P, van der Windt DA, van der Horst HE, et al. Prediction of an unfavourable course of low back pain in general practice: comparison of four instruments. *Br J Gen Pract*. 2007;57:15–22.
12. 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;14:83–89.
13. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159–174.
14. Main CJ, Watson P. Assessment and management of the distressed and angry low back pain patients in primary care, private practice and community settings. In: Gifford L, ed. *Topical Issues in Pain. Vol. 3*. Falmouth: CNS Press; 2001:175–200.
15. Ong BN, Hooper H. Comparing clinical and lay accounts of the diagnosis and treatment of back pain. *Sociol Health Illn*. 2006;28:203–222.
16. Dunn KM, Jordan K, Croft PR. Recall of medication use, self-care activities and pain intensity: a comparison of daily diaries and self-report questionnaires among low back pain patients. *Prim Health Care Res Dev*. 2010;11:93–102.

© 2010 Lippincott Williams & Wilkins