The Spine Journal 19 (2019) 645–654

Clinical Study

The use of STarT back screening tool to predict functional disability outcomes in patients receiving physical therapy for low back pain

Irene L. Katzan, MD[a,*], Nicolas R. Thompson, MS[a], Steven Z. George, PT, PhD, FAPTA[b], Sandi Passek, PT, DPT[c], Frederick Frost, MD[c], Mary Stilphen, PT, DPT[c]

[a] *Neurological Institute Center for Outcomes Research & Evaluation, Neurological Institute, Cleveland Clinic, 9500 Euclid Avenue, Cleveland Ohio 44195, USA*

[b] *Duke Clinical Research Institute and Department of Orthopaedic Surgery, Duke University, 2400 Pratt Street, Room 0311 Terrace Level, Durham NC 27705, USA*

[c] *Department of Physical Medicine & Rehabilitation, Neurological Institute, Cleveland Clinic, 9500 Euclid Avenue, Cleveland Ohio 44195, USA*

Received 22 July 2018; revised 2 October 2018; accepted 2 October 2018

Abstract

BACKGROUND CONTEXT: The STarT Back Screening Tool (SBST) categorizes risk of future disability in patients with low back pain (LBP). Previous studies evaluating the use of SBST in physical therapy (PT) populations do not reflect the ethnic and socioeconomic diversity occurring in clinical practice and lack statistical power to evaluate factors associated with outcomes within each SBST risk category.

PURPOSE: The purpose of this study is to further refine SBST risk categorization for predicting improvements in functional disability with attention toward patient level factors that might guide SBST use in routine outpatient physical therapy practice.

STUDY DESIGN/SETTING: This was a retrospective cohort study that took place within a large academic, tertiary-care health system.

PATIENT SAMPLE: The study cohort consisted of 1,169 patients with LBP who completed a course of outpatient physical therapy from June 1, 2014 to May 31, 2015 and who completed the patient-reported SBST and modified low back pain disability questionnaire (MDQ) questionnaires as part of standard of care.

OUTCOME MEASURES: Improvement in functional disability defined as decrease in 10 or more points in the MDQ.

METHODS: Multivariable logistic regression was performed to evaluate independent predictors of improvement after PT, which included SBST risk category, baseline MDQ, a two-way interaction term between SBST category and baseline MDQ, prior level of function (independent vs. required assistance), demographic characteristics, number of completed PT visits, and duration of PT episode of care. In exploratory analyses, additional two-way interaction terms between SBST category and the significant predictors were added to the regression model.

RESULTS: Mean age of patients in the study cohort was 55.1 years (SD 16.1); 657 (56.2%) were female, 117 (10.0%) were black race, 127 (10.9%) had Medicaid insurance, and 353 (30.2%) had previously received PT for back pain. In all, 35.8% (n=419) patients categorized as low risk SBST category, 40.7% (n=476) medium risk SBST category, and 23.4% (n=274) high risk SBST category. There was an interaction between baseline MDQ and SBST risk category and improvement

FDA device/drug status: Not applicable.

Author disclosures: ***IK***: Nothing to disclose. ***NT***: Nothing to disclose. ***SG***: Nothing to disclose. ***SP***: Nothing to disclose. ***FF***: Nothing to disclose. ***MS***: Nothing to disclose.

* Corresponding author. Neurological Institute Center for Outcomes Research & Evaluation, Cleveland Clinic, 9500 Euclid Ave, S80, Cleveland, OH 44195, USA. Tel.: +12164452616.

E-mail address: katzani@ccf.org (I.L. Katzan).

https://doi.org/10.1016/j.spinee.2018.10.002

with PT. For all three SBST categories, higher baseline MDQ was associated with higher probability of improvement, but the effect was less pronounced as SBST risk category increased. Additional factors independently associated with reduced odds of improvement after PT included black race (odds ratio [OR] 0.44, 95% confidence interval [CI] 0.28–0.72), Medicaid insurance (OR=0.58, 95% CI 0.36–0.95), and prior PT (OR=0.48, 95% CI 0.34–0.67). In exploratory analyses, there was a significant interaction between insurance type and SBST risk category in predicting functional improvement after PT. Patients with Medicare and Medicaid insurance had similar rates of improvement in low and high risk SBST categories but different rates of improvement in the medium risk categories.

CONCLUSIONS: The SBST tool predicts outcomes of PT in a cohort of patients receiving outpatient PT for LBP. The odds of improvement varied according to baseline disability and SBST risk status. Race, insurance type, and history of previous PT influenced prediction independent of SBST risk status. Incorporating these variables and the interaction between SBST and baseline disability in outcome models has the potential to refine prediction of outcomes after PT.

Keywords: Low back pain; Outcomes; Physical therapy; Prediction; STarT back screening tool; Modified low back pain disability questionnaire.

Introduction

Back pain is one of the most common health problems with more than two-thirds of the population reporting back pain at least once during life [1,2]. Low back problems are the second and third leading complaint in outpatient visits and hospital admissions respectively [1,3] and carry an annual prevalence of as high as 45% [4]. The high incidence and prevalence of back pain equates to substantial health care utilization and spending. In 2005 in the US, the direct health care costs of back pain were estimated at 85 billion dollars [5,6]. More concerning is that these cost has increased by 65% since 1995 at a rate far outpacing overall healthcare spending [5–7].

At least 25% of outpatient physical therapy practice is related to low back pain (LBP) [8–10]. The American College of Physicians practice guidelines recommend many nonpharmacologic interventions used by physical therapists as front line LBP treatments [11]. However, 45% of patients receiving physical therapy fail to have significant improvement in pain and disability [12]. Early identification of those with favorable and nonfavorable prognoses could have a significant impact on future care models for LBP. In particular, emotional dimensions are important contributors to outcomes of patients with LBP. The STarT Back Screening Tool (SBST) [13] is one well-established clinical tool used to determine risk of future disability based on psychosocial factors. It was originally designed for use in primary care [14], where providing risk-stratified treatment improved disability outcomes while lowering healthcare utilization when compared to usual care [15]. The SBST has also been used to successfully predict disability outcomes for patients seeking outpatient physical therapy [16,17].

There are still areas to explore for gaining a better understanding of the SBSTs predictive capabilities, in order to improve their utility in clinical practice. Previous studies have focused on differences in clinical outcomes between the three risk groups, because they lacked adequate statistical power to refine predictive accuracy within an individual risk group. Prior investigations were recruited as part of clinical trials or cohort studies, and may not reflect ethnic and socioeconomic diversity occurring in real practice settings. Future investigation of these factors is needed as highlighted by a recent analysis indicating individuals with greater psychosocial distress and lower socioeconomic status (SES) had decreased benefit of receiving risk stratified care [18]. Therefore, the overall purpose of this study to further refine SBST risk categorization for predicting improvements in disability in a multivariable model with attention toward patient level factors that might guide SBST use in routine outpatient physical therapy practice. First, we determined the association of SBST categorization with meeting a clinically meaningful improvement threshold for functional disability, as determined by the patient-reported modified low back pain disability questionnaire (MDQ). Second, we investigated whether there was an interaction between SBST risk category and baseline functional disability to provide better context in predicting improvement. Information from this analysis will optimize the use of SBST within clinical practice.

Methods

Study design

This was a retrospective cohort study of patients with LBP who presented for outpatient physical therapy at Cleveland Clinic Health System facilities in northeastern Ohio. In this system, patient-reported outcomes are systematically collected at the point of care on tablets at all physical therapy sites through the Knowledge Program data collection system [19]. Results are immediately available within the electronic health record (Epic, Epic Corporation,

Verona, WI, USA). SBST and MDQ were both collected during routine clinical encounters. Data from the first year of collection, June 1, 2014 to May 31, 2015, were included in this analysis.

All physical therapists follow the Cleveland Clinic Physical Therapy Back Pain Care Path, which incorporates evidence-based principles of PT, standardized documentation, and outcomes assessment for patients with back and neck pain (see Appendix I). A treatment-based classification system is utilized, which matches patients treatments to examination findings [20–24] and may include corrective exercises for motor control, back specific strength and conditioning, directional preference and movement retraining, and manual therapy.

Data elements

Prognostic measure

The SBST consists of 9 items, which are used to categorize patients' level of risk for having persisting LBP with disability: low risk if the total score from both subscales is from 0 to 3, high risk if the psychosocial subscale score is 4 or 5, and medium risk if falling into neither the low-risk nor the high-risk [13]. The SBST was used as a categorical variable (low, medium, or high risk) for data analysis.

Outcome measure

The MDQ is a validated 10-item questionnaire instrument [25] based on the Oswestry Disability Index. It assesses different aspects of function in LBP. Each item is scored 0 to 5 points, which are summed and doubled to produce a final score ranging from 0 to 100. Higher scores indicate greater disability. Reduction in MDQ of 10 or more points has been demonstrated to indicate clinically meaningful improvement [26–28] and indicates a moderate or greater effect size [11]. This threshold was used in our analysis to indicate a clinically meaningful improvement (10+ point change) or not (less than 10 point change) during a given treatment episode.

Covariates

All demographic, clinical, and treatment variables were electronically extracted from the electronic health record. Demographic variables included age (in years), sex (Male, Female), race (White, Black, Other), ethnicity (Hispanic vs. non-Hispanic), marital status (married, single, divorced, widowed), and insurance provider (private, self-pay, Medicare, Medicaid). Clinical variables were documented by physical therapists in standardized templates at the initial visit of the defined episode of care and included number of recommended visits, working status (working, not working, retired, disabled), previous treatments for LBP (physical therapy, chiropractic treatment, massage, acupuncture, injections, exercise, and surgery). In addition, prior function before episode of LBP was documented. Patients were categorized as "required assistance" if they required help with ambulation, activities of daily living, independent activities of daily living, or required workplace accommodation. Otherwise patients were categorized as "independent." Treatment variables included completed PT visits (number of visits) and duration of PT episode of care (number of days).

Patient selection

Patients were included in this analysis if they were over the age of 18 years at their first visit in the study period and during their first PT episode of care had a primary International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) code for LBP (724.2, 724.3, 724.4, and 724.5). Consistent with the date range included in the data analysis, patients' first visit in the episode of care had to occur on or after June 1, 2014 and their last visit attended in the episode of care had to be on or before May 31, 2015. Furthermore, in order to be included in the analysis patients must have completed the SBST and MDQ at baseline and the MDQ at a follow-up visit. For patients with more than one follow-up MDQ, we used the last one completed for a given episode of care. Patients were excluded if there was inadequate clinical documentation to identify when the PT episode of care was completed.

Data analysis

Clinical characteristics of patients in the study sample were summarized using descriptive statistics for patients in the study sample overall and stratified by SBST category. Baseline, follow-up, and change in MDQ score were computed in each of the SBST categories and we used analysis of variance to determine whether the mean change in score was different among the three groups. In addition, cumulative distribution functions were graphed to examine the proportion of patients in each SBST risk category who experienced change in MDQ score along a continuum of change from improvement to worsening. A log-rank test was performed to determine if the cumulative distribution functions were different for the three SBST risk categories.

Multivariable logistic regression

Multivariable logistic regression was performed to evaluate the independent effect of SBST risk category on improvement in functional disability with PT after adjustment for clinical and demographic characteristics and PT treatment variables. Continuous covariates were treated as linear and categorical variables were categorized as described in the above covariates section. Improvement in MDQ score of 10 or more points was the response variable. The independent variable of interest was the baseline SBST category. We also examined whether the effect of SBST category on improvement in MDQ depended on baseline MDQ. This was achieved by including two-way interaction terms between SBST category and these variables. In exploratory analyses, additional interaction terms between

SBST and all other factors found to be independently associated with improvement were then included in the model. C-statistics for each model were computed and adjusted for optimism using bootstrap internal validation methods. Model calibration was assessed graphically by plotting predicted probabilities versus observed probabilities.

All computations were done in R, version 3.4.1 [29]. All tests were two-sided and p values less than .05 were considered statistically significant. Missing covariate data (eg, race, work status) was handled using multivariate imputation by chained equations. The imputation model was created in the full cohort of back pain patients and subsequent outcome models created on the subset of patients who completed the baseline SBST and MDQ at baseline and follow-up [30]. This study was approved by the Cleveland Clinic Institutional Review Board. Because data used in this study was collected as part of standard of care, requirement for informed consent was waived.

Results

Patient characteristics

A total of 10,329 patients had a PT episode of care between June 1, 2014 and May 31, 2015 with a primary ICD-9-CM diagnosis code for a LBP and were 18 years of age or older at their first visit. We removed 287 patients who were missing documentation indicating their episode of care was complete and an additional 6,130 patients who did not complete the SBST. Finally, we removed 2,743 patients who did not have an MDQ at baseline and follow-up. Thus, our final sample consisted of 1,169 patients (Fig. 1).

Descriptive statistics for patients included and excluded from analysis data are presented in Appendix II, Table A1. Patients included in the study cohort were slightly more likely to be female, more likely to be white, had slight differences in insurance status, were less likely to have previously required assistance, more likely to have had previous chiropractic treatment, and had a larger average number of recommended PT visits and larger average visits attended.

In patients in the analyzed cohort, 35.8% (n=419) were categorized as low risk (SBST Category 1), 40.7% (n=476) medium risk (SBST Category 2), and 23.4% (n=274) high risk (SBST Category 3). Follow-up MDQ score was completed a mean of 44.6 days (standard deviation [SD] 29.3) after baseline MDQ score. Patients in the high risk SBST category were younger, more likely to be black, less likely to be married, lived in zip codes with lower median income, were less likely to be working, were more likely to be on Medicaid, were more likely to have previously required assistance, were less likely to have used acupuncture in the past, had higher baseline MDQ, and had a larger average number of recommended visits as well as visits attended (Table 1).

As expected, higher baseline SBST category was associated with higher MDQ score at baseline and follow-up (Table 2). There were also differences in the distributions of MDQ change scores for the 3 groups ($p < .001$), with greater improvement in patients in the medium and high risk SBST categories (Fig. 2).

Independent predictors of functional improvement

Of the 1,169 patients included in the multivariable logistic regression model, 537 (45.9%) improved by 10 or more points. There were 4 variables independently associated with improvement in MDQ. Black patients had lower probability of improvement compared to white patients (odds ratio [OR] 0.44, 95% confidence interval [CI] 0.28–0.72). Patients on Medicaid were less likely to improve than patients with private health insurance (OR=0.58, 95% CI 0.36–0.95). Patients who had prior PT were less likely to report meaningful improvement; none of the other six

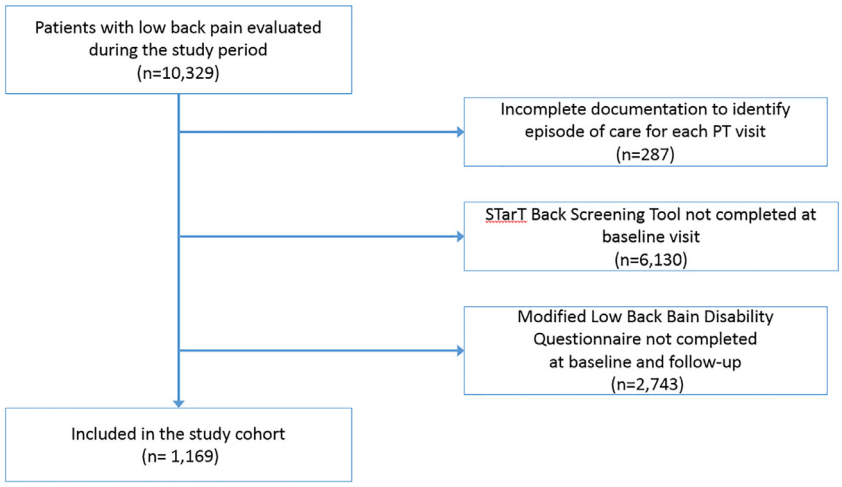

Fig. 1. Low back pain population.
Flow diagram of study subjects.

Table 1
Descriptive statistics of patient characteristics stratified by SBST risk category

	All patients	SBST risk category=1	SBST risk category=2	SBST risk category=3	p value
N	1,169	419	476	274	
Age, mean (SD)	55.1 (16.1)	57.1 (16.6)	54.2 (15.7)	53.4 (15.6)	.004
Sex					
Male	512 (43.8%)	191 (45.6%)	195 (41.0%)	126 (46.0%)	.269
Female	657 (56.2%)	228 (54.4%)	281 (59.0%)	148 (54.0%)	
Race					
White	973 (83.2%)	365 (87.1%)	403 (84.7%)	205 (74.8%)	< .001
Black	117 (10.0%)	26 (6.2%)	45 (9.5%)	46 (16.8%)	
Other	29 (2.5%)	14 (3.3%)	6 (1.3%)	9 (3.3%)	
Missing	50 (4.3%)	14 (3.3%)	22 (4.6%)	14 (5.1%)	
Ethnicity					
Hispanic	38 (3.3%)	12 (2.9%)	11 (2.3%)	15 (5.5%)	.056
Non-Hispanic	1,107 (94.7%)	401 (95.7%)	452 (95.0%)	254 (92.7%)	
Missing	24 (2.1%)	6 (1.4%)	13 (2.7%)	5 (1.8%)	
Marital status					
Married	670 (57.3%)	258 (61.6%)	288 (60.5%)	124 (45.3%)	.002
Single	296 (25.3%)	99 (23.6%)	114 (23.9%)	83 (30.3%)	
Divorced	100 (8.6%)	34 (8.1%)	35 (7.4%)	31 (11.3%)	
Widowed	75 (6.4%)	19 (4.5%)	32 (6.7%)	24 (8.8%)	
Missing	28 (2.4%)	9 (2.1%)	7 (1.5%)	12 (4.4%)	
Median income by ZIP code (x $1,000), mean (SD)	55.4 (18.2)	58.3 (18.4)	54.7 (17.6)	52.0 (18.4)	< .001
Work status					
Working	682 (58.3%)	249 (59.4%)	293 (61.6%)	140 (51.1%)	< 0.001
Not working	115 (9.8%)	24 (5.7%)	39 (8.2%)	52 (19.0%)	
Retired	293 (25.1%)	135 (32.2%)	115 (24.2%)	43 (15.7%)	
Disabled	54 (4.6%)	6 (1.4%)	19 (4.0%)	29 (10.6%)	
Missing	25 (2.1%)	5 (1.2%)	10 (2.1%)	10 (3.6%)	
Insurance					
Private	555 (47.5%)	197 (47.0%)	248 (52.1%)	110 (40.1%)	< .001
Self-pay	195 (16.7%)	80 (19.1%)	76 (16.0%)	39 (14.2%)	
Medicare	249 (21.3%)	104 (24.8%)	84 (17.6%)	61 (22.3%)	
Medicaid	127 (10.9%)	24 (5.7%)	49 (10.3%)	54 (19.7%)	
Missing	43 (3.7%)	14 (3.3%)	19 (4.0%)	10 (3.6%)	
Prior function					
Independent	1095 (93.7%)	407 (97.1%)	446 (93.7%)	242 (88.3%)	< .001
Require assistance	39 (3.3%)	3 (0.7%)	16 (3.4%)	20 (7.3%)	
Missing	35 (3.0%)	9 (2.1%)	14 (2.9%)	12 (4.4%)	
Previous treatments					
Physical therapy	353 (30.2%)	132 (31.5%)	134 (28.2%)	87 (31.8%)	.450
Occupational therapy	2 (0.2%)	0 (0.0%)	1 (0.2%)	1 (0.4%)	.708
Chiropractic	160 (13.7%)	47 (11.2%)	70 (14.7%)	43 (15.7%)	.172
Massage	36 (3.1%)	9 (2.1%)	16 (3.4%)	11 (4.0%)	.342
Acupuncture	15 (1.3%)	3 (0.7%)	11 (2.3%)	1 (0.4%)	.031
Injections	121 (10.4%)	37 (8.8%)	52 (10.9%)	32 (11.7%)	.420
Exercise	232 (19.8%)	86 (20.5%)	82 (17.2%)	64 (23.4%)	.117
Surgery	35 (3.0%)	17 (4.1%)	11 (2.3%)	7 (2.6%)	.276
Baseline MDQ score, mean (SD)	33.8 (17.3)	20.7 (11.9)	37.1 (14.3)	48.0 (15.1)	< .001
Number of recommended visits, mean (SD)	8.6 (4.2)	7.9 (4.2)	8.7 (3.9)	9.5 (4.3)	< .001
Number of visits attended, mean (SD)	7.1 (3.7)	6.6 (3.6)	7.3 (3.7)	7.6 (3.9)	< .001
Duration of episode of care (days), median (IQR)	42 (29, 63)	42 (29, 62)	42 (28, 63)	44.5 (30, 69.5)	.690

IQR, interquartile range; MDQ, modified low back pain disability questionnaire; SBST, STarT Back Screening Tool; SD, standard deviation. SBST category score indicates risk for long-term disability: 1, low risk, 2, medium risk, and 3, high risk.

previous treatments were independently associated with outcome (Table 3).

There was an interaction between baseline MDQ and SBST risk category in the logistic regression model examining the factors associated with improvement in MDQ by 10 or more points. For all three SBST categories, higher baseline MDQ was associated with higher probability of improving by 10 or more points in MDQ, but the effect was less pronounced as SBST risk category increased (see Fig. 3).

Model discrimination was moderate (optimism-adjusted c-statistic=0.695, based on 200 bootstrap iterations). A calibration plot revealed that predicted probabilities below 0.1

Table 2
Means and standard deviations of modified low back pain disability questionnaires at baseline, follow-up, and the change in score, stratified by SBST risk category

	SBST risk category=1 N=419	SBST risk category=2 N=476	SBST risk category=3 N=274	p value
Mean MDQ score (SD)				
Baseline	20.7 (11.9)	37.1 (14.3)	48.0 (15.1)	<.001
Follow-up	13.7 (11.7)	25.1 (17.1)	37.1 (20.2)	<.001
Change*	−7.0 (11.2)	−12.0 (15.6)	−10.9 (18.0)	<.001
Clinically important change in MDQ†, n (%)				
Improved	162 (38.7%)	248 (52.1%)	127 (46.4%)	Improved
Stable	235 (56.1%)	198 (41.6%)	123 (44.9%)	Stable
Worsened	22 (5.3%)	30 (6.3%)	24 (8.8%)	Worsened

* Change is defined as follow-up score − baseline score.
† Clinically important change defined as change in MDQ ≥ 10 points.

were slightly underestimated while predicted probabilities above 0.9 were slightly overestimated (see Appendix II, Figure A1). Although, more than 96% of predicted probabilities fell between 0.1 and 0.9, indicating that model calibration was good for the majority of our sample.

To estimate the probability of improvement by 10 points in MDQ for a particular patient, one can input the patient's information into the model using the estimated coefficients (not ORs) found in Table 3. The resulting value on the log-odds scale can be converted to a predicted probability by employing the inverse logit function (see Appendix III).

Exploratory analysis

Two-way interaction terms between SBST risk category and the variables race, insurance, and previous PT were added to the model in exploratory analyses. The interaction term between insurance and SBST risk category was statistically significant (omnibus p value=.013). There was a variable pattern in the proportion of patients who improved by 10 points in the MDQ across SBST categories according to insurance status. Patients with Medicare and Medicaid insurance had similar rates of improvement in low and high risk SBST categories but different rates of improvement in the medium risk categories. Only patients with private insurance showed a consistent trend in outcomes across SBST risk categories, with lower rates of improvement with increasing risk category. Interactions between race and SBST risk category and prior PT and SBST category were not significant (omnibus p values=.688 and .585, respectively). Model discrimination was very similar to the model without the additional interaction terms (optimism-adjusted c-statistic=0.697) as was calibration (see Appendix II, Figure A2).

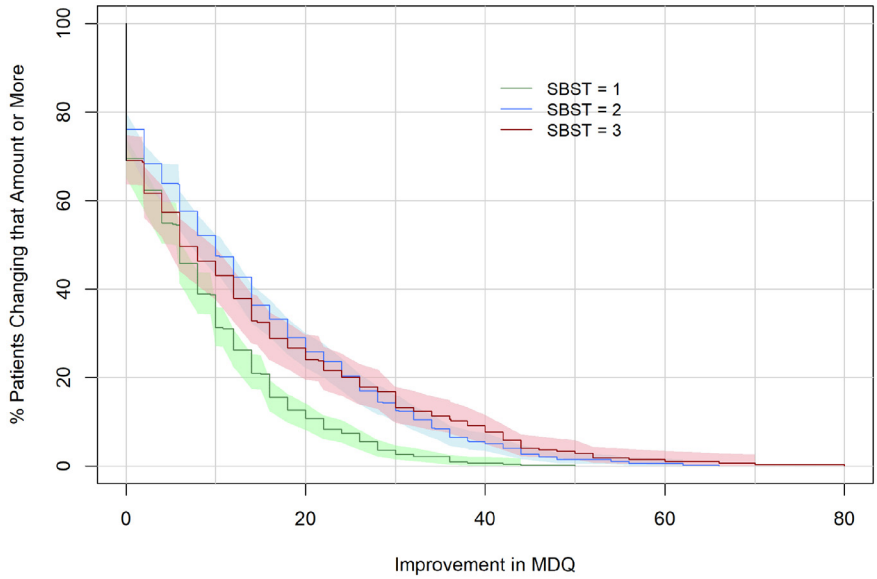

Fig. 2. Cumulative frequency distributions stratified by SBST risk category.
MDQ, modified low back pain disability questionnaire; SBST, STarT Back Screening Tool. This graph demonstrates the change in MDQ on the x −axis and the percent of patients experiencing that change on the Y-axis stratified by baseline SBST score. As an example of how to interpret this graph, approximately 15% of patients with SBST score of 1 and 30% of patients with SBST score of 2 or 3 demonstrated improvement of 20 points or greater in the MDQ.

Table 3
Results of multivariable logistic regression model where dependent variable was improvement in MDQ by 10 or more points

Covariate	Odds ratio (95% CI)	p value
Intercept	0.16 (0.06, 0.41)	< .001
SBST risk category (vs. 1)		
SBST 2	2.63 (1.22, 5.68)	.014
SBST 3	4.89 (1.79, 13.37)	.002
Baseline MDQ score (per 10 units)	2.59 (2.06, 3.25)	< .001
Age (per 10 y)	0.96 (0.85, 1.07)	.450
Sex (vs. female)		
Male	1.02 (0.79, 1.33)	.862
Race (vs. white)		
Black	0.44 (0.28, 0.72)	< .001
Other	1.08 (0.46, 2.56)	.855
Ethnicity (vs. non-Hispanic)		
Hispanic	0.52 (0.22, 1.23)	.135
Marital status		
Single	0.80 (0.57, 1.12)	.197
Divorced	0.99 (0.61, 1.60)	.966
Widowed	1.21 (0.69, 2.12)	.498
Median household income* (per $10,000)	1.04 (0.96, 1.13)	.300
Insurance (vs. private)		
Self-pay	0.88 (0.61, 1.27)	.500
Medicare	0.96 (0.64, 1.44)	.842
Medicaid	0.58 (0.36, 0.95)	.030
Work status (vs. working)		
Not working	0.69 (0.42, 1.11)	.123
Retired	0.90 (0.59, 1.36)	.612
Disabled	0.61 (0.31, 1.18)	.142
Prior function (vs. independent)		
Require assistance	0.74 (0.34, 1.61)	.452
Previous physical therapy	0.48 (0.34, 0.67)	< .001
Previous chiropractic treatment	1.05 (0.73, 1.53)	.781
Previous massage	0.70 (0.34, 1.44)	.338
Previous acupuncture	0.90 (0.29, 2.75)	.851
Previous injections	0.68 (0.44, 1.04)	.072
Previous exercise	1.07 (0.74, 1.55)	.731
Previous surgery	0.78 (0.36, 1.68)	.527
Number of physical therapy visits	1.00 (0.95, 1.04)	.819
Duration of episode of care (per 30 d)	0.92 (0.78, 1.08)	.299
Baseline MDQ×SBST (vs. SBST 1)†		
SBST 2	0.61 (0.47, 0.80)	< .001
SBST 3	0.48 (0.36, 0.64)	< .001

CI, confidence interval; PT, physical therapy.
* Median household income estimated from ZIP code.
† See Fig. 3 for graphical depiction of interaction between baseline MDQ and SBST.

Discussion

Our analyses demonstrated a potentially complex relationship between SBST category, disability score, and improvement. The interaction between SBST category and baseline MDQ score is a novel finding with clinical relevance, as our findings indicated that the relationship between SBST category and meaningful improvement in disability varied according to baseline MDQ scores. Among patients with lower baseline disability (MDQ scores < 20), those categorized as low psychosocial risk had lower probabilities of improvement than patients with medium or high psychosocial risk. Conversely, in patients with higher baseline disability (MDQ scores > 20), patients categorized as low risk had the greatest probability of improvement with progressively higher disability scores, followed by patients with medium psychosocial risk. Patients characterized as high risk had the lowest probability of improvement as the baseline MDQ scores increased. These data are some of the first we are aware of that show how (1) accuracy of SBST risk stratification could be refined by using the baseline disability, and (2) other prognostic factors used in conjunction with the SBST allow better prediction of clinical improvements.

The lower effectiveness of PT in patients in the high risk SBST category with MDQ scores > 20 suggests that patients with moderate or greater disability scores and psychological distress may benefit from interventions that supplement traditional PT and that includes a focus on psychological aspects of back-pain related disability. This interpretation is supported by the success of a stratified approach to back pain care in primary care and outpatient physical therapy settings [15,31]. The greater improvement seen in patients with low SBST category compared to the medium and high risk categories may reflect the favorable natural history of LBP regardless of treatment in this patient group [13]. Because we did not have a comparator group of patients who did not receive PT, we were not able to determine which explanation is correct.

In addition to the SBST interaction with baseline MDQ score, Medicaid insurance, black race, and previous PT were each independently associated with lower odds of improvement in disability. Medicaid insurance is often considered a surrogate for low SES [32], which has been previously associated with lower response to intervention for back pain SBST [18]. However, median household income estimated from ZIP code, another marker for SES [33], was not significantly associated with response to treatment in our cohort. It is possible that income estimated from ZIP code is inadequate to assess SES in this context, as SES is a multidimensional construct that includes education level and occupation in addition to income [34]. Alternatively, Medicaid insurance in this analysis may reflect poor health more than SES. Poor health has been consistently associated with worse prognosis in patients with back pain [35]. Of note, there was a significant interaction between insurance and SBST in the prediction model, with variable pattern of improvement across SBST categories according to insurance type. These patterns are difficult to explain but suggest that patients with different insurance types differ in ways that were unmeasured in this analysis but which impact the outcomes of patients receiving PT.

Racial differences in the experience of pain are well-recognized. Black patients with pain have greater pain-related symptoms and disability compared to white patients [36–38] and report significantly more post-traumatic stress disorder, irritability, and depressive symptoms than older non-Hispanic whites [37]. Less is known about the impact of race on response to treatment of pain although the

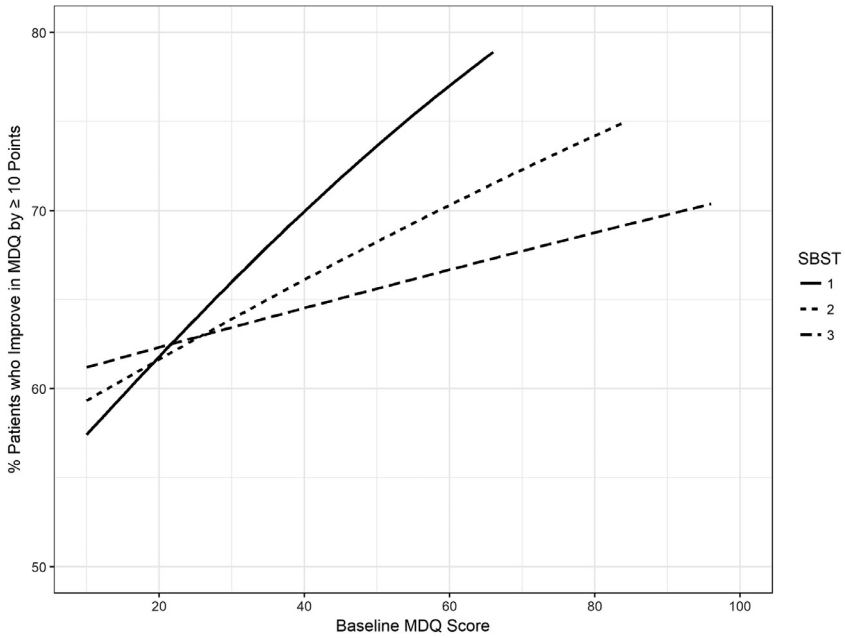

Fig. 3. Predicted percentage of patients who improved by 10 or more points on the MDQ by SBST risk category and baseline MDQ.

MDQ, modified low back pain disability questionnaire; SBST, STarT Back Screening Tool. Figure values adjusted for variables in Table 3. Predictions were estimated by inputting median values for continuous variables and reference categories for categorical variables.

limited available data have shown similar treatment benefit between black and white patients [39,40]. We found no interaction between race and SBST risk category and improvement after PT indicating that the predictive value of SBST in Black patients is similar across risk categories. There is a need for further research to clarify factors contributing to the association of race with worse outcomes with PT demonstrated in our study and how these factors can be combined with SBST to refine outcome prediction.

The reduced adjusted odds for clinically significant improvement in patients who underwent previous PT is, by itself, not surprising; these patients were more likely to have had recurrent or chronic LBP and therefore poorer prognosis. However, there was no association between improvement and the other six previous treatments for LBP, which would be expected if prior PT served only as a surrogate for chronic LBP. These findings raise the question of whether patients who have undergone previous PT should be considered for supplemental or alternate treatment for LBP, possibly with interventions that addresses psychosocial factors contributing to LBP. These findings could also be used to generate hypotheses for future studies of how prior PT can be used to inform prediction of clinical improvements.

Collectively these findings indicate that prior PT, race, and insurance add supplemental information to outcome prediction models that include the interaction between baseline SBST and MDQ score. These two scales incorporate many of the constructs relevant for prognostication in patients with LBP [41]. The clinical relevance of these findings is the demonstration of how prediction of functional outcomes can be refined without adding items to the SBST. One of the inherent advantages of the SBST clinical use is its brevity but there are still opportunities for improving its accuracy for outcome prediction. These analyses incorporated information from routine clinical encounters and identified how it can be used to improve outcome prediction in tandem with the SBST. In clinical settings where there are options to increase item sets additional constructs not included in our dataset related to resilience and health behaviors may further improve the ability to predict outcomes [41]. Further research is needed to determine whether including variables that measure these constructs add incremental prognostic information to SBST and functional disability for use in individualizing treatment and whether they mediate the relationship between the variables race, prior PT, and insurance status on disability outcomes.

An important strength of this study is the large sample size collected during routine clinical visits. This allowed us to determine the generalizability of SBST risk status beyond clinical trial and cohort studies, as well as more closely assess the relationship between SBST and clinical variables. This sample also allowed to determine the probabilities of improvement with varying baseline levels of functional disability within each SBST category, which provides information that could be incorporated into treatment decisions. In addition, our study cohort was more racially diverse than the patient populations in the previous SBST studies [42], allowing us to evaluate the association between race and outcomes after PT.

This study also has several limitations. The study cohort consisted of patients who had 1 of 4 primary encounter diagnosis codes for nonspecific back pain and the specific

pain triggers and anatomic abnormalities likely varied across patients. This approach may have introduced unmeasured variability in the analyses by not directly accounting for specific diagnostic factors. However, this approach is defensible because it increased generalizability of our study results and there is little evidence to support diagnostic categories being strongly linked to outcomes [43,44]. Patient included in the study cohort met specific inclusion criteria, and represented 10% of patients with 1 of 4 encounter diagnosis codes, raising the possibility of selection bias. Although we corrected for measured variables in regression models, it is possible that study patients are different from excluded patients in ways that were unmeasured. To account for selection bias using the available data, our imputation model was created in the full cohort of back pain patients and subsequent outcome models created on the subset of patients who completed the MDQ [30]. Another limitation is that the same minimal clinically important difference for the MDQ was used for all patients, which may limit responsiveness depending on LBP duration and severity. In addition, the PT program followed basic principles of therapy with individualization of content and number of recommended sessions based on the therapist's initial evaluation (see Appendix I). This may have reduced our ability to detect differences in outcomes according to MDQ score and SBST level. However, individualizing PT is a basic tenant of care; this analysis is intended to determine if additional prognostic factors may be helpful at optimizing the prediction of clinical outcomes. The duration of LBP was not available in this analysis, although the use of previous PT, a potential surrogate for duration, was included. Finally, another limitation is that a myriad of other variables that may have impacted outcomes, such as specific type of PT used (eg, spinal manipulation), adherence to home exercise, adherence to recommended number of therapy visits, and comorbid conditions, were not included in this analysis.

Conclusion

This analysis has shown that SBST tool predicts outcomes of PT in a real world cohort of patients receiving outpatient physical therapy for LBP. We identified nuances that may refine future SBST use. First, there was an interaction indicating that odds of improvement varied according to baseline disability and SBST risk status. Second, we identified race, insurance type, and history of previous PT as variables that improved prediction independent of SBST risk status. Incorporating these variables and the interaction between SBST and baseline disability in outcome models has the potential to refine prediction of outcomes after PT. Collectively these findings could have implications for clinical management by identifying nonmodifiable social (ie, race, insurance type) and utilization (ie, prior PT) factors that are not captured by the SBST. The complex relationship between SBST and these other factors merit further investigation in future studies.

Acknowledgment

Dr. George acknowledges funding from the National Institutes of Health/National Center for Complementary and Integrative Health (UG3AT009790) in preparation of this manuscript.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.spinee.2018.10.002.

References

[1] Deyo RA, Weinstein JN. Low back pain. N Engl J Med 2001;344:363–70.
[2] Biyani A, Andersson GB. Low back pain: pathophysiology and management. J Am Acad Orthop Surg 2004;12:106–15.
[3] Irving GA, Wallace MS. Pain management for the practicing physician. New York: Churchill Livingstone; 1997.
[4] Andersson GB. Epidemiological features of chronic low-back pain. Lancet 1999;354:581–5.
[5] Martin BI, Turner JA, Mirza SK, Lee MJ, Comstock BA, Deyo RA. Trends in health care expenditures, utilization, and health status among US adults with spine problems, 1997-2006. Spine 2009;34:2077–84.
[6] Freburger JK, Holmes GM, Agans RP, Jackman AM, Darter JD, Wallace AS, et al. The rising prevalence of chronic low back pain. Arch Intern Med 2009;169:251–8.
[7] Stewart WF, Ricci JA, Chee E, Morganstein D, Lipton R. Lost productive time and cost due to common pain conditions in the US workforce. JAMA 2003;290:2443–54.
[8] Di Fabio RP, Boissonnault W. Physical therapy and health-related outcomes for patients with common orthopaedic diagnoses. J Orthop Sports Phys Ther 1998;27:219–30.
[9] Jette AM, Delitto A. Physical therapy treatment choices for musculoskeletal impairments. Phys Ther 1997;77:145–54.
[10] Resnik L, Liu D, Mor V, Hart DL. Predictors of physical therapy clinic performance in the treatment of patients with low back pain syndromes. Phys Ther 2008;88:989–1004.
[11] Qaseem A, Wilt TJ, McLean RM, Forciea MA. Clinical Guidelines Committee of the American College of P: noninvasive treatments for acute, subacute, and chronic low back pain: a clinical practice guideline from the American College of Physicians. Ann Intern Med 2017;166:514–30.
[12] Heneweer H, Aufdemkampe G, van Tulder MW, Kiers H, Stappaerts KH, Vanhees L. Psychosocial variables in patients with (sub)acute low back pain: an inception cohort in primary care physical therapy in the Netherlands. Spine 2007;32:586–92.
[13] Hill JC, Dunn KM, Lewis M, Mullis R, Main CJ, Foster NE, et al. A primary care back pain screening tool: identifying patient subgroups for initial treatment. Arthritis Rheum 2008;59:632–41.
[14] Suri P, Delaney K, Rundell SD, Cherkin DC. 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.
[15] Hill JC, Whitehurst DG, Lewis M, Bryan S, Dunn KM, Foster NE, 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.
[16] Beneciuk JM, Fritz JM, George SZ. The STarT Back Screening Tool for prediction of 6-month clinical outcomes: relevance of change

patterns in outpatient physical therapy settings. J Orthop Sports Phys Ther 2014;44:656–64.
[17] Beneciuk JM, Bishop MD, Fritz JM, Robinson ME, Asal NR, Nisenzon AN, 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.
[18] Beneciuk JM, Hill JC, Campbell P, Afolabi E, George SZ, Dunn KM, et al. Identifying treatment effect modifiers in the STarT Back Trial: a secondary analysis. J Pain 2017;18:54–65.
[19] Katzan I, Speck M, Dopler C, Urchek J, Bielawski K, Dunphy C, et al. The Knowledge Program: an innovative, comprehensive electronic data capture system and warehouse. AMIA Annu Symp Proc 2011;2011:683–92.
[20] Kendall N, Linton SJ, Main CJ. Guide to assessing psychosocial yellow flags in acute low back pain: risk factors for long-term disability and work loss. Wellington, New Zealand: Accident Compensation Corporation and the New Zealand Guidelines Group,: Wellington; 2004.
[21] Nijs J, Van Houdenhove B. From acute musculoskeletal pain to chronic widespread pain and fibromyalgia: application of pain neurophysiology in manual therapy practice. Man Ther 2009;14:3–12.
[22] Nijs J, Van Houdenhove B, Oostendorp RA. Recognition of central sensitization in patients with musculoskeletal pain: application of pain neurophysiology in manual therapy practice. Man Ther 2010;15:135–41.
[23] Smart KM, Blake C, Staines A, Thacker M, Doody C. Mechanisms-based classifications of musculoskeletal pain: part 1 of 3: symptoms and signs of central sensitisation in patients with low back (+/− leg) pain. Man Ther 2012;17:336–44.
[24] Kent P, Kjaer P. The efficacy of targeted interventions for modifiable psychosocial risk factors of persistent nonspecific low back pain - a systematic review. Man Ther 2012;17:385–401.
[25] Fritz JM, Irrgang JJ. A comparison of a modified Oswestry Low Back Pain Disability Questionnaire and the Quebec Back Pain Disability Scale. Phys Ther 2001;81:776–88.
[26] Vianin M. Psychometric properties and clinical usefulness of the Oswestry Disability Index. J Chiropr Med 2008;7:161–3.
[27] Davidson M, Keating J. Oswestry Disability Questionnaire (ODQ). Aust J Physiother 2005;51:270.
[28] Davidson M, Keating JL. A comparison of five low back disability questionnaires: reliability and responsiveness. Phys Ther 2002;82:8–24.
[29] The R Core Team. R: a language and environment for statistical computing. www.R-project.org. In R Foundation for Statistical Computing: Vienna, Austria; 2017.
[30] von Hippel PT. Regression with missing Ys: an improved strategy for analyzing multiply imputed data. Sociol Methodol 2007;37:83–117.
[31] Beneciuk JM, George SZ. Pragmatic implementation of a stratified primary care model for low back pain management in outpatient physical therapy settings: two-phase, sequential preliminary study. Phys Ther 2015;95:1120–34.
[32] Marcin JP, Schembri MS, He J, Romano PS. A population-based analysis of socioeconomic status and insurance status and their relationship with pediatric trauma hospitalization and mortality rates. Am J Public Health 2003;93:461–6.
[33] Berkowitz SA, Traore CY, Singer DE, Atlas SJ. Evaluating area-based socioeconomic status indicators for monitoring disparities within health care systems: results from a primary care network. Health Serv Res 2015;50:398–417.
[34] American Psychological Association. Report of the APA task force on socioeconomic status. Availiable at http://www.apa.org/pi/ses/resources/publications/index.aspx; 2007.
[35] Hayden JA, Chou R, Hogg-Johnson S, Bombardier C. Systematic reviews of low back pain prognosis had variable methods and results: guidance for future prognosis reviews. J Clin Epidemiol 2009;62:781–96.
[36] Riley JL, 3rd, Wade JB, Myers CD, Sheffield D, Papas RK, Price DD. Racial/ethnic differences in the experience of chronic pain. Pain 2002;100:291–8.
[37] Green CR, Baker TA, Sato Y, Washington TL, Smith EM. Race and chronic pain: a comparative study of young black and white Americans presenting for management. J Pain 2003;4:176–83.
[38] Campbell CM, Edwards RR. Ethnic differences in pain and pain management. Pain Manag 2012;2:219–30.
[39] Gagnon CM, Matsuura JT, Smith CC, Stanos SP. Ethnicity and interdisciplinary pain treatment. Pain Pract 2014;14:532–40.
[40] Merry B, Campbell CM, Buenaver LF, McGuire L, Haythornthwaite JA, Doleys DM, et al. Ethnic group differences in the outcomes of multidisciplinary pain treatment. J Musculoskelet Pain 2011;19:24–30.
[41] Boissoneault J, Mundt J, Robinson M, George SZ. Predicting low back pain outcomes: suggestions for future directions. J Orthop Sports Phys Ther 2017;47:588–92.
[42] Fritz JM, Beneciuk JM, George SZ. Relationship between categorization with the STarT Back Screening Tool and prognosis for people receiving physical therapy for low back pain. Phys Ther 2011;91:722–32.
[43] Hartvigsen J, Hancock MJ, Kongsted A, Louw Q, Ferreira ML, Genevay S, et al. What low back pain is and why we need to pay attention. Lancet 2018;391:2356–67.
[44] Foster NE, Anema JR, Cherkin D, Chou R, Cohen SP, Gross DP, et al. Prevention and treatment of low back pain: evidence, challenges, and promising directions. Lancet 2018;391:2368–83.